

Contactless Skin Temperature Management for Preterm Infants: A Multi-Modal Approach

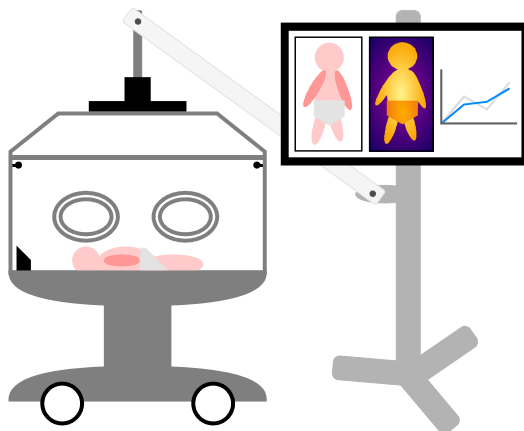
Von der Fakultät für Elektrotechnik und Informationstechnik der
Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des
akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte
Dissertation

vorgelegt von

Florian Voss, M.Sc.
aus Haan, Deutschland

Berichter: Univ.-Prof. Dr.-Ing. Dr. med. Dr. h.c. Steffen Leonhardt
Univ.-Prof. Dr. med. Thorsten Orlikowsky

Tag der mündlichen Prüfung: 12.11.2025



Contactless Skin Temperature Management for Preterm Infants: A Multi-Modal Approach

Von der Fakultät für Elektrotechnik und Informationstechnik der
Rheinisch-Westfälischen Technischen Hochschule Aachen zur Erlangung des
akademischen Grades eines Doktors der Ingenieurwissenschaften genehmigte
Dissertation

vorgelegt von

Florian Voss, M.Sc.
aus Haan, Deutschland

Berichter: Univ.-Prof. Dr.-Ing. Dr. med. Dr. h.c. Steffen Leonhardt
Univ.-Prof. Dr. med. Thorsten Orlikowsky

Tag der mündlichen Prüfung: 12.11.2025

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

Acknowledgments

This dissertation is the result of my work at the Chair of Medical Information Technology (MedIT) at RWTH Aachen University. I would therefore like to express my special thanks to my supervisor, Prof. Steffen Leonhardt. I am grateful for his scientific guidance, his support, and the freedom he gave me to pursue my own ideas, which I believe is the most important driving force of science.

I would also like to thank my co-examiner, Prof. Thorsten Orlikowsky, for his time and effort and for his valuable feedback.

My special gratitude goes to my direct supervisors, Prof. Christoph Hoog Antink and Dr. Markus Lüken. They supervised my daily work, always encouraged my ideas, and were available for scientific advice. Their constructive feedback and guidance fundamentally shaped my approach to research and scientific writing.

This work would not have been possible without constructive collaborations. I would like to thank my project partners, especially Dr. Konrad Heimann, Dr. Milian Brasche, Jörg Weyer, and Jan Ingenhag, for their support, expertise, and the excellent teamwork.

Several students contributed to the research described in this thesis as part of their bachelor's and master's theses. I would like to thank all of them for their hard work and dedication.

This work would not have been nearly as enjoyable without my colleagues at MedIT. I thank the entire group for the great work atmosphere and helpful discussions. I will especially cherish the memories we made outside of research: our intense games of Kicker, making music with the MedIT Band, and almost winning the Institute Olympiad together twice. These moments balanced the intense five-year journey of my Ph.D. well.

Words cannot express my gratitude to my family and friends. Thank you for your encouragement, understanding, and for providing a healthy and necessary balance to my work. I thank my parents for their extensive support in my life and for always believing in me.

A final and deepest thank you goes to my wife, Luisa, for your love, your patience, and for supporting me in this journey, from the moment we met. I am incredibly grateful to have you by my side.

Abstract

Maintaining stable skin temperature is critical for the survival and healthy development of premature infants. Traditional methods that rely on wired skin sensors carry the risk of skin damage, infection and interference with parent-infant interaction.

This thesis presents a novel, non-contact approach using Infrared Thermography to address these challenges, enabling continuous and accurate temperature monitoring without physical contact. For this purpose, an automated body part segmentation algorithm has been developed as a critical first step for accurate temperature extraction from infrared images. A dataset of multimodal RGB and long-wave infrared data was created and annotated. Deep learning models, specifically U-Net architectures, were trained and evaluated, demonstrating that multi-modal fusion approaches combining information from both RGB and LWIR images significantly outperform single-modality methods in segmenting body regions such as the head, torso, arms and legs.

Accurate IRT-based temperature measurement within an incubator requires careful consideration of environmental influences such as IR window transmission and reflections. This work introduces a novel temperature correction algorithm that addresses these challenges and has been calibrated using a black body. This algorithm significantly reduced measurement errors, achieving a mean absolute error of 0.17°C compared to the black body. A clinical study of premature infants validated the performance of the combined segmentation and temperature correction pipeline, demonstrating its accuracy and robustness in a real-world environment. In the clinical setting, the system achieved a mean absolute error of 0.41°C compared to adhesive skin sensors. Furthermore, the torso and arms were found to be the most reliable regions for contactless temperature measurement due to their lower fluctuations and higher detectability.

A novel neonatal thermal phantom was created to enable the development and validation of non-contact skin temperature control algorithms. This phantom, which incorporates a 3D-printed anatomical structure with independently controllable heating zones and embedded temperature sensors, accurately simulates the thermal behavior of a premature infant, allowing realistic testing of various physiological and pathological scenarios. Finally, a hardware-in-the-loop (HIL) test bed was designed, integrating the IRT system, segmentation models and the thermal phantom. PI-based control strategies were implemented and evaluated, demonstrating successful closed-loop control of infant skin temperature by regulating the incubator air temperature, while minimizing fluctuations and responding effectively to simulated perturbations and impaired thermoregulation.

This research provides a foundation for the future development of fully contactless closed-loop skin temperature control systems, promising improved safety and comfort for premature infants.

Zusammenfassung

Die Aufrechterhaltung einer stabilen Hauttemperatur ist für das Überleben und die gesunde Entwicklung von Frühgeborenen von entscheidender Bedeutung. Herkömmliche Methoden, die auf kabelgebundenen Hautsensoren basieren, bergen das Risiko von Hautschäden, Infektionen und einer Beeinträchtigung der Eltern-Kind-Interaktion. Diese Arbeit stellt einen neuartigen, kontaktlosen Ansatz mittels Infrarot-Thermografie vor, um diese Herausforderungen zu bewältigen und eine kontinuierliche und genaue Temperaturüberwachung ohne physischen Kontakt zu ermöglichen.

Zu diesem Zweck wurde ein automatisierter Algorithmus zur Körperteilsegmentierung als kritischer erster Schritt für die genaue Temperaturextraktion aus Infrarotbildern entwickelt. Es wurde ein Datensatz multimodaler RGB- und Langwellen-Infrarotbildern erstellt und annotiert. Deep-Learning-Modelle, insbesondere U-Net-Architekturen, wurden trainiert und evaluiert. Es zeigte sich, dass multimodale Fusionsansätze, die Informationen aus RGB- und LWIR-Bildern kombinieren, bessere Segmentierungsergebnisse liefern als Modelle mit nur einer Bildmodalität.

Eine genaue IRT-basierte Temperaturmessung in einem Inkubator erfordert die sorgfältige Berücksichtigung von Umgebungseinflüssen wie Störungen durch das IR-Fenster und Reflexionen. Diese Arbeit stellt einen neuartigen Temperaturkorrekturalgorithmus vor, der diese Herausforderungen adressiert und mithilfe eines Schwarzkörpers kalibriert wurde. Dieser Algorithmus reduzierte den Messfehler signifikant und erreichte einen mittleren absoluten Fehler (Mean Absolute Error) von 0.17°C im Vergleich zum Schwarzkörper. Eine klinische Studie an Frühgeborenen validierte die Kombination aus Segmentierungs- und Temperaturkorrektur-Pipeline und demonstrierte deren Genauigkeit und Robustheit in einer realen Umgebung. Im klinischen Umfeld erreichte das System einen mittleren absoluten Fehler von 0.41°C im Vergleich zu klebenden Hautsensoren. Darüber hinaus erwiesen sich der Torso und die Arme aufgrund ihrer geringeren Schwankungen und höheren Detektierbarkeit als die zuverlässigsten Regionen für die kontaktlose Temperaturmessung.

Ein neuartiges neonatales thermisches Phantom wurde entwickelt, um die Entwicklung und Validierung von Algorithmen zur kontaktlosen Hauttemperaturregelung zu ermöglichen. Dieses Phantom, das eine 3D-gedruckte anatomische Struktur mit unabhängig steuerbaren Heizzonen und eingebetteten Temperatursensoren umfasst, simuliert das thermische Verhalten eines Frühgeborenen präzise und ermöglicht realistische Tests verschiedener physiologischer und pathologischer Szenarien. Schließlich wurde ein Hardware-in-the-Loop (HIL) Prüfstand konzipiert, der das IRT-System, die Segmentierungsmodelle und das thermische Phantom integriert. PI-basierte Regelungsstrategien wurden

implementiert und evaluiert. Sie zeigten eine erfolgreiche Regelung (Closed-Loop-Control) der Hauttemperatur des Säuglings durch die Regulierung der Inkubator-Lufttemperatur, wobei Schwankungen minimiert und effektiv auf simulierte Störungen sowie eine beeinträchtigte Thermoregulation reagiert wurde.

Diese Forschung legt den Grundstein für die zukünftige Entwicklung vollständig kontaktloser, geschlossener Regelungssysteme für die Hauttemperatur, die eine verbesserte Sicherheit und einen höheren Komfort für Frühgeborene versprechen.

Contents

Acknowledgments	v
Abstract	vii
Zusammenfassung	ix
Abbreviations and symbols	xv
1 Introduction	1
2 Medical Background and Technical Fundamentals	5
2.1 Medical Background	5
2.1.1 Premature Birth	5
2.1.2 Thermoregulation of Premature Infants	7
2.1.3 Neonatal Incubators	13
2.2 Infrared Thermography	15
2.2.1 Infrared Radiation	16
2.2.2 Infrared Sensors and Cameras	17
2.2.3 Absolute Temperature Measurement	19
2.3 Deep Learning	21
2.3.1 Artificial Neural Networks	22
2.3.2 Convolutional Neural Networks	23
2.3.3 Semantic Segmentation	24
2.3.4 Training and Overfitting	26
2.4 Control Theory	29
2.4.1 System Identification	30
2.4.2 Uncertainty Modeling	31
2.4.3 Hardware-in-the-Loop	32
3 Body Part Segmentation of Premature Infants using a Multi-Modal Camera System	35
3.1 Motivation and Overview	35
3.2 Materials and Methods	37
3.2.1 Dataset	37
3.2.2 Network Architecture	40
3.2.3 Data Augmentation and Transfer Learning	44
3.2.4 Evaluation Method	45
3.3 Results	49
3.3.1 Hyperparameter Optimization	49

3.3.2	Evaluation on Test Set	51
3.4	Discussion	54
3.4.1	Hyperparameter Tuning	54
3.4.2	Image Fusion	56
3.4.3	Class Prediction	57
3.5	Summary	58
4	Spatial Temperature Monitoring of Premature Infants	61
4.1	Motivation and Overview	61
4.2	Clinical Study	63
4.2.1	Hardware Setup	63
4.2.2	Recordings	65
4.3	Accurate Thermography for Incubators	66
4.3.1	Setup for Calibration	68
4.3.2	Calibration Algorithm	69
4.3.3	Evaluation Measurements	72
4.3.4	Validation Results and Discussion	73
4.4	Body Part Segmentation	75
4.4.1	Dataset	75
4.4.2	Segmentation Models	78
4.4.3	Evaluation Method	79
4.4.4	Results and Discussion	80
4.5	Spatial Temperature Extraction	88
4.5.1	Methods	89
4.5.2	Results and Discussion	91
4.6	Summary	97
5	A Physical Phantom for Neonatal Thermoregulation	99
5.1	Motivation and Overview	99
5.2	Materials and Methods	100
5.2.1	3D-Modeling	100
5.2.2	Thermoregulation	102
5.2.3	Thermal Control	104
5.2.4	Optimization of Thermal Homogeneity	107
5.3	Results and Discussion	112
5.3.1	Validation of temperature control	113
5.3.2	Skin Temperature Control	115
5.3.3	Pathology Simulation	117
5.4	Summary	118

6	Contactless Skin Servo Control for Neonatal Incubators	121
6.1	Motivation and Overview	121
6.2	Hardware-in-the-Loop Setup	123
6.2.1	Automated Temperature Measurement	124
6.2.2	Neonatal Phantom	125
6.2.3	Software of the Test Bench	128
6.3	Control System	128
6.3.1	System Modeling	129
6.3.2	System Identification	131
6.3.3	Control Design	134
6.4	Results	136
6.4.1	Disturbance Rejection	136
6.4.2	Impaired Thermoregulation	137
6.5	Discussion	143
6.6	Summary	145
7	Conclusion	147
A	Appendix	151
A.1	Backbone Testing	151
A.2	Bode Plots	152
B	Publications	153
	Literaturverzeichnis	157
	Curriculum Vitae	175

Abbreviations and symbols

List of abbreviations

Abbrev.	Meaning
ADC	Analog-to-digital converters
BAT	Brown adipose tissue
CIHP	Crowd Instance-level Human Parsing
CLAHE	Contrast Limited Adaptive Histogram Equalisation
CNN	Convolutional neural network
cpTD	Central-peripheral temperature difference
DenseNet	Densely Connected Convolutional Networks
FOV	Field of view
HIL	Hardware-in-the-loop
IRT	Infrared thermography
IoU	Intersection-over-Union
LBW	Low birth weight
LTI	linear time-invariant
LWIR	Long-wave infrared
mIoU	mean Intersection-over-Union
MWIR	Mid-wave infrared
MSE	Mean square error
movSD	Moving standard deviation
MWIR	Mid-wave infrared
NICU	Neonatal intensive care unit
NIR	Near infrared
PCB	Printed circuit board
PI	Proportional integral
PWM	Pulse width modulation
RANSAC	Random Sample Consensus
ReLU	Rectified Linear Unit
ResNet	Residual Neural Network
RGB	Red, green, blue
ROI	Region of interest
RMSE	Root Mean Square Error

Abbrev.	Meaning
SDG	Sustainable Development Goal
SIL	Software-in-the-Loop
SNR	Signal-to-noise ratio
SPR	Set point range
STD	Standard deviation
UART	Universal Asynchronous Receiver/Transmitter
UKA	RWTH Aachen University Hospital
VLBW	Very low birth weight
WHO	World Health Organization

List of Latin Symbols

Symbol	Unit	Meaning
$A_{\text{Skin,cond.}}$	m^2	Conductively effective skin surface
$A_{\text{Skin,conv.}}$	m^2	Convectively effective skin surface
$A_{\text{Skin,rad.}}$	m^2	Radiative effective skin surface
E_{obj}	W/m^2	Emission intensity by a target object
E_{surr}	W/m^2	Reflected emission intensity by ambient objects/surroundings
E_{atm}	W/m^2	Emission from the atmosphere
c	m/s	Speed of light
c_{infant}	J/kgK	Specific heat capacity of the infant
$e(t)$		Control error of control system
f_{var}		Heat factor
G_0		Open control loop transfer function
G_C		Controller transfer function
$G_{C,\text{Air}}$		Air temperature controller transfer function
$G_{C,\text{Skin}}$		Skin temperature controller transfer function
$G_{P,\text{Air}}$		Air temperature plant transfer function
$G_{P,\text{Incubator}}$		Combined Incubator plant transfer function
$G_{P,\text{Incubator, down}}$		Incubator plant transfer function (decreasing temperature)
$G_{P,\text{Incubator, up}}$		Incubator plant transfer function (increasing temperature)
$G_{P,\text{Infant,lin}}$		Linearized infant plant transfer function
$G_{P, \text{Skin}, \nearrow}$		Time-variant infant plant transfer function
$G_{S(s)}$		Transfer functions of the controlled system
h_c	$\text{W}/\text{m}^2\text{K}$	Convective heat transfer coefficient
h	Js	Planck's constant
k_b	J/K	Boltzmann's constant
K_c	W/mK	Heat conduction coefficient
K_i		Integral gain of the PI controller
$K_{\text{Incubator, down}}$		Incubator plant gain (decreasing temperature)
$K_{\text{Incubator, up}}$		Incubator plant gain (increasing temperature)
K_g		Static gain of a system

Symbol	Unit	Meaning
K_p		Proportional gain of the PI controller
m_{infant}	kg	Mass of the infant
$\dot{Q}_{\text{Conduction}}$	W	Heat loss due to conduction
$\dot{Q}_{\text{Convection}}$	W	Heat loss due to convection
Q_{Kalman}	°C/s	Process noise of the Kalman filter
$\dot{Q}_{\text{metabolic}}$	W	Metabolic heat production
$\dot{Q}_{\text{metabolic, var}}$	W	Time-variant metabolic heat production
$\dot{Q}_{\text{Radiation}}$	W	Heat loss due to radiation
$\dot{Q}_{\text{total}}$	W	Total emitted radiation intensity
$\dot{Q}_{\text{heat}}$	W	Total heat production of the infant phantom
$\dot{Q}_{\text{heat, var}}$	W	Variable heat production of the infant phantom
R'	Ω/m	Resistance per unit length
R_{Kalman}	°C	Measurement noise of the Kalman filter
t	s	Time
T	K	Temperature
T_{Air}	K	Ambient air temperature
T_{amb}	K	Temperature of the ambient sources
T_{atm}	K	Temperature of the atmosphere
$T_{\text{BB, Corr}}$	K	Corrected temperature of the blackbody
$T_{\text{BB, IRT}}$	K	IRT-measured temperature of the blackbody
$T_{\text{BB, real}}$	K	True temperature of the blackbody
T_{cam}	K	Camera temperature
$T_{\text{cam, corr}}$	K	NUC corrected IRT temperature
$T_{\text{hom., cam}}$	K	Thermal image of calibration surface
T_{Inc}	°C	Air temperature of the incubator
$T_{\text{Inc, real}}$	°C	Real air temperature of the incubator
$T_{\text{Inc, set}}$	°C	Set point air temperature of the incubator
$T_{\text{Inc, sim}}$	°C	Simulated incubator air temperature
$T_{\text{Incubator, down}}$	s	Incubator time constant (decreasing temperature)
$T_{\text{Incubator, up}}$	s	Incubator time constant (increasing temperature)
T_{inf}	K	Infant's skin temperature
T_{Mattress}	K	Temperature of the mattress
$T_{\text{Nuc, Mat}}$	K	Non-uniformity correction matrix
T_{obj}	K	Temperature of a target object

Symbol	Unit	Meaning
T_p		Pole point of a system
$T_{\text{Ref,IRT}}$	K	IRT-measured temperature of a thermal reference
$T_{\text{Ref,Sens}}$	K	Sensor-based temperature of a thermal reference
$T_{\text{Ref,Surf}}$	K	True temperature of a calibration surface
T_{Surr}	K	Temperature of surrounding objects
T_{Skin}	K	Skin temperature
$T_{\text{Skin,real}}$	°C	Real skin temperature
$T_{\text{Skin,set}}$	°C	Set point skin temperature
$T_{\text{Skin, sim}}$	°C	Simulated infant skin temperature
T_t		Dead time of a system
T_z		Zero point of a system
$u(t)$		Manipulated variable of control system
$w(t)$		Reference variable of control system
W_{cam}	W	Thermal radiation emitted by the camera
W_{inf}	W	Infant's emitted radiation
W_{surr}	W	Thermal radiation emitted by incubator surroundings
W_{tot}	W/m ²	Total radiation detected by a IRT camera
W_{win}	W	Thermal radiation emitted by IR window
$y(t)$		Controlled variable of control system
$z(t)$		Disturbance of control system

List of Non-Latin Symbols

Symbol	Unit	Meaning
α_{RD}	°	Phase reserve
ΔT	s	Sampling time
ϵ		Emission coefficient
ϵ_{inf}		Infant's skin emissivity
ϵ_{skin}		Emission coefficient of the skin
ϵ_{atm}		Emission coefficient of the atmosphere
ϵ_{obj}		Emission coefficient of a target object
ϵ_{win}		Emission coefficient of IR window
τ		Transmission coefficient
τ_{atm}		Transmission coefficient of atmosphere
τ_{win}		Transmission coefficient of IR window
ϕ	°	Phase margin
ρ		Reflection coefficient
ρ_{obj}		Reflection coefficient of a target object
ρ_{win}		Reflection coefficient of IR window
λ	m	Wavelength
λ_{max}	m	Wavelength of the intensity maximum
σ	W/m ² K ⁴	Stefan-Boltzmann constant
ω	rad/s	Angular frequency

1 Introduction

One of the biggest challenges of newborn care is the risk of preterm birth. When an infant arrives prematurely, defined as birth before 37 weeks of gestation, it presents significant challenges for both the infant and the family. The World Health Organization (WHO) states that the first 28 days of life carry the highest risk of dying during infancy [Wor20]. The mortality rate in the neonatal period might be up to 30 times higher [LBJH13]. Even though technical progress and improved medical care have already led to a 50 % reduction in child mortality from 1990 until 2019 [Eur11], the WHO estimates that still 2.4 million newborns were lost in 2019 which means an average of about 6700 neonatal deaths per day worldwide [Wor20], [UUW⁺20]. Neonatal deaths account for more than 40 % of under-5 mortality, making preterm birth the leading global under-5 cause of neonatal death according to recent estimations (about 1.055 million deaths), even surpassing pneumonia (about 0.921 million), and intrapartum-related events (0.691 million) [LBJH13]. This is due to the fact, that the reduction in mortality caused by the latter reasons progressed faster compared to prematurity [LOH⁺16]. Despite the high mortality rate, only 2.5 % of global data on neonatal deaths are based on reliable registration systems, while the rest are based on estimates and surveys [LBJH13]. There is still a lack of knowledge in this field.

In addition, the rates of preterm birth are rising. In 2012, the WHO reported that depending on the financial situation, 5-18 % of all newborns (about 15 million) are born prematurely. The mortality of a premature infant strongly decreases with higher gestational age, higher birthweight and better, more expensive care [Wor12]. The latter causes a lot more premature infants to be lost in low-income countries than in high-income countries [Wor12], [Wor18].

One of the most critical aspects of neonatal care for premature infants is thermoregulation. Unlike full-term infants, premature infants have a limited capacity to maintain their body temperature within the optimal range [LBJH13]. This vulnerability stems from several factors. Firstly, their larger surface area to body mass ratio leads to faster heat loss to the environment. Secondly, they have less subcutaneous and brown fat, which provides insulation and helps generate heat [KGR11], [CWH⁺05]. Additionally, their skin is not fully developed, resulting in increased water loss through evaporation and less effective blood flow control. Furthermore, their immature heart, lung, and muscle systems also contribute to their compromised thermoregulation abilities [LSC19], [WM07].

Consequently, premature infants are highly susceptible to cold stress, which can lead to a cascade of adverse outcomes. When the body temperature falls below the optimal range, the infant's metabolism increases to generate heat, leading to increased oxygen demand and decreased production of surfactant, a

substance crucial for lung function. This can result in low oxygen and blood sugar levels, hindering growth and development. Additionally, cold stress can contribute to apnea and even death [LP15].

To address this vulnerability, neonatal intensive care units (NICUs) rely on incubators to provide a controlled microclimate that supports the infant's thermoregulation. Monitoring an infant's thermal condition can also help to detect early signs of abnormal clinical states such as cold or heat stress as well as infections by closely tracking not only the central but also the peripheral temperature [LP15], [KHDSW09].

However, current monitoring commonly relies on wired thermistors attached to the infant's skin. These sensors can cause skin damage, irritation and increase the risk of infection, particularly in premature infants with fragile and underdeveloped skin [ODR15]. Additionally, the wires can hinder parent-infant bonding, complicate medical procedures, and restrict the infant's movement within the incubator [BBCL17].

The limitations of contact-based skin temperature monitoring methods have led to the research of non-contact alternatives. Infrared thermography (IRT) is emerging as a promising solution, using the ability to measure the infrared radiation emitted by the skin to determine temperature. This technology offers several distinct advantages over traditional methods. Firstly, IRT is non-invasive and does not require physical contact with the infant's skin, reducing the risk of skin damage and infection associated with wired sensors. Secondly, IRT allows continuous temperature monitoring without disturbing the infant, which could offer valuable insight into their thermal stability and response to interventions. Finally, IRT provides spatially resolved temperature information that maps the heat distribution across the infant's body. This comprehensive perspective could provide a deeper understanding of their thermoregulation compared to conventional single-point measurements.

These advantages make IRT a highly attractive option for improving the temperature management and comfort of premature infants. However, its clinical application requires automated segmentation of infant skin within the IRT image data to account for movement artifacts that render manual segmentation impractical for continuous monitoring. In addition, the accuracy of infrared cameras currently is insufficient compared to contact-based sensors, which limits their clinical use and requires improvement.

Aim of the Thesis

This thesis presents a novel approach to contactless skin temperature monitoring in premature infants, addressing the limitations of contact-based methods.

It develops and validates an IRT-based system for both thermal monitoring and skin temperature control, enabling robust and accurate contactless skin temperature management. Specifically, the thesis addresses the following topics:

- Algorithms are presented to automatically segment infant body parts using multimodal imaging data, consisting of visible (red, green, blue - RGB) and long-wave infrared (LWIR) images, to accurately identify and segment different body regions such as the head, torso, arms and legs. By fusing information from both modalities, a more robust and accurate segmentation is introduced.
- A spatial skin temperature measurement system using IRT and deep learning based segmentation algorithms is presented that is able to accurately measure the temperature distribution in different body regions without physical contact with the infant. This eliminates the risks associated with skin-attached sensors, improves the comfort and well-being of the infant, and potentially reduces the workload of healthcare staff.
- Control algorithms for contactless skin servo control in neonatal incubators are introduced using a novel hardware-in-the-loop test bench. The development of control strategies is presented that uses the contactless temperature sensing system to adjust the incubator air temperature and maintain the infant's skin temperature at a desired set point.

By achieving these objectives, this thesis aims to improve the care and comfort of premature infants and provide a foundation for further research and clinical translation of non-contact skin temperature management methods.

Structure of the Thesis

This thesis is divided into seven chapters to provide a comprehensive overview of the conducted research.

Chapter 2 presents the foundation by providing the necessary medical background on preterm birth and infant thermoregulation. It explains the specific challenges faced by premature infants and the importance of maintaining optimal thermal conditions for their development and survival. In addition, this chapter introduces the technical fundamentals of IRT, deep learning and control theory, which are essential for understanding the proposed contactless approach.

Chapter 3 focuses on the development and evaluation of deep learning models for infant body part segmentation. It compares different deep learning approaches with a U-Net architecture and explores different fusion approaches

for combining RGB and LWIR image data. This chapter investigates the effectiveness of transfer learning and data augmentation techniques to improve the accuracy and robustness of the segmentation models.

Chapter 4 describes the implementation of the contactless skin temperature measurement system and its validation in a clinical study. This includes the hardware setup with RGB and LWIR cameras, the development of temperature correction algorithms to ensure accurate IRT measurements and the integration of the deep learning model for real-time skin segmentation. This chapter presents the validation of the system on real data and discusses its potential for further clinical application.

Chapter 5 introduces a novel neonatal phantom designed to simulate the thermoregulation of a premature infant. This phantom is equipped with heating elements and sensors to mimic the thermal behavior of a real infant. It allows testing and validation of the non-contact temperature measurement system and control algorithms in various physiological and pathological scenarios.

Chapter 6 presents the development and evaluation of contactless skin servo control for neonatal incubators using a novel hardware-in-the-loop (HIL) setup. This setup integrates the contactless temperature measurement system and the neonatal phantom to create a realistic test environment. This chapter discusses the design and implementation of the HIL setup and presents initial controller designs for contactless skin servo control. The performance of these controllers is evaluated through simulation and real-world testing on the HIL test bench.

Chapter 7 concludes the thesis by summarizing the key findings, discussing limitations, and outlining potential directions for future research. This chapter highlights the potential impact of this research on improving the care and comfort of premature infants and emphasizes the need for further development and clinical validation.

2 Medical Background and Technical Fundamentals

The purpose of this chapter is to lay the groundwork for the subsequent chapters. It consists of three main sections. The first section (Section 2.1) presents the medical context, explaining the causes and consequences of premature birth (Section 2.1.1), the thermoregulation challenges faced by premature infants (Section 2.1.2) and the essential aspects of neonatal care (Section 2.1.3). The second section (Section 2.2) describes the basics of infrared thermography, describing its physical background. The third section (Section 2.3) briefly introduces Deep Learning and its key concepts, while the last section (Section 2.4) introduces the necessary concepts of control theory.

2.1 Medical Background

This section introduces the medical background relevant to the subsequent chapters. It covers the basics of preterm birth, the thermoregulation of premature infants and the fundamentals of neonatal care. These aspects are essential for understanding the physiological and environmental factors that affect the development and survival of premature infants.

2.1.1 Premature Birth

Prematurity is a significant cause of death, disability and long-term complications for newborns worldwide, affecting not only the health and well-being of the infants and their families but also creating an intergenerational health problem with lasting social and economic consequences.

According to the WHO [Wor12], about 15 million infants are born premature every year worldwide, which means between 5 % and 18 % of all live births. Moreover, premature birth rates are rising in almost all countries with reliable data, partly due to changes in maternal age, lifestyle and reproductive technologies. Premature birth or prematurity refers to infants who are born alive before completing the 37th week of pregnancy. There are different subcategories of premature birth based on the gestational age, which is the period from the first day of the last menstrual cycle until birth. The majority of premature births (approximately 80 %) are of a moderate to late gestational age, occurring between 32 and 37 weeks, and have a relatively favourable prognosis. The remaining 20 % of premature births are very premature (28 to 32 weeks of gestation) and extremely premature (less than 28 weeks of gestation), in decreasing order [Wor12], [Wor18]. These infants face higher risks of mortality and morbidity and require more intensive care.

Another important classification is based on the absolute birth weight or the birth weight relative to the average birth weight for a given gestational age. An infant is considered to have low birth weight (LBW) if it weighs less than 2500 g at birth or very low birth weight (VLBW) if it weighs less than 1500 g. These infants are more vulnerable to infections, hypothermia and other complications. On average, premature infants have a hospital stay that is ten times longer and health care costs that are nine times higher than term infants. If possible, they spend most of their time in neonatal intensive care units (NICUs), which provide optimal conditions for them regarding temperature and humidity. These factors are crucial for maintaining their body temperature and preventing heat loss. However, two-thirds of all premature births occur in low-income settings that usually lack the resources and equipment to provide the necessary care for their survival [Wor12], [LBJH13], [Eur19b].

The exact causes of premature birth are still unclear and the risk factors change with the development of prevention strategies, so in most cases, no specific cause is found. However, some factors that increase the likelihood of premature birth are known, such as maternal stress, age extremes, multiple pregnancies, previous premature births, infections and chronic conditions. Most premature births (two-thirds) occur spontaneously, while others are medically or non-medically induced [Wor12], [Wor18], [ZAB⁺18]. Premature birth is not a local problem but a global challenge. About 60 % of all premature infants are born in South Asia and sub-Saharan Africa, where some countries have the highest rates of premature births compared to all live births (up to 18 % in Malawi). Unfortunately, two-thirds of all premature births happen in low-income settings that usually lack the resources and equipment to provide the necessary care for their survival. High-income and middle-income countries can afford dedicated intensive care units for premature infants. In contrast, often poorer countries can barely provide a clean, fully equipped facility with enough staff and space [Eur19a], [Wor12].

This unequal distribution of infant healthcare situations also results in a higher mortality rate in low-income countries. While high-income countries manage to save up to 90 % of extremely premature infants (< 28 weeks), only 10 % of infants of the same gestation survive in low-income countries. On average, this still means a mortality of about 6.6 %. Low birth weight and a young gestational age strongly increase the chance of dying, making premature birth the leading cause of death among children under five years [Wor12], [Wor18], [Wor20], [ZAB⁺18].

In [SMTK⁺08], Shapiro et al. found that the newborn morbidity risk increased twofold in each gestational week earlier than 38 weeks and that late-premature infants had a seven-times higher risk of morbidity compared to term-born infants (22 % vs. 3 %). Although overall under-5 mortality decreased

significantly, neonatal mortality is declining at a much slower rate (see Figure 2.1). Besides the immediate risk of death, premature-born infants face a higher risk of developing long-term health problems such as non-communicable diseases, hypertension, diabetes and generally chronic diseases and mortality later in life [Wor12], [ZAB⁺18].

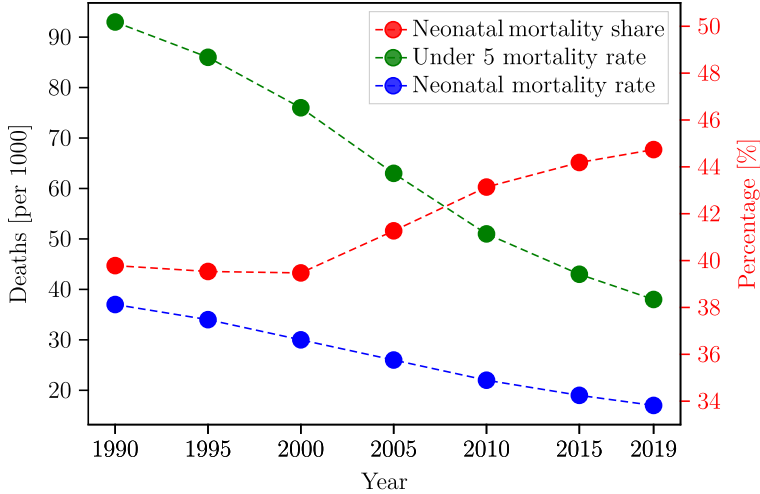


Abbildung 2.1: Development of neonatal mortality compared to under-5 mortality over the last 30 years (values from [UW⁺20]).

Furthermore, the United Nations’ 2030 Agenda for Sustainable Development (SDGs) sets the goal to lower neonatal mortality to at most 1.2 % and under-5 mortality to 2.5 % [Uni21]. To achieve this goal, the WHO published an action report [Wor12], [Wor18] claiming that most premature infants can be saved with feasible, cost-effective care. Their suggestions include general, essential care for mothers and infants, healthy pregnancy guidelines, antenatal supportive measures and postnatal best practices and they especially recommend kangaroo care and midwife-led care services. Moreover, collecting and providing reliable data is essential to support and form evidence-based decisions.

2.1.2 Thermoregulation of Premature Infants

Premature infants have a limited capacity for thermoregulation and require assistance to maintain their body temperature within the normo-thermal range, as abnormal body temperature is associated with adverse outcomes, morbidity

and mortality in the NICU [LSX⁺15], [SAU13], [LBJH13]. In order to optimize the support with technical means, it is essential to comprehend the inherent thermoregulation of an infant [LP15]. The term *thermoregulation* denotes the primary function of preserving the body temperature in the optimal range regardless of variations in environmental temperature by modulating heat loss and heat production via feedback control [JPDE⁺17]. It can be conceptualized as a control loop with the body temperature as the set or target variable, the body as the process or controlled system that is influenced by environmental disturbance variables that several actuators attempt to counterbalance with the brain as the controller that accesses the body temperature via a feedback line.

The optimal range of body temperature for humans is around 37 °C for both adults [SBP09] and infants [LP15], [LSC19]. Typical mean body temperatures 15 min after birth are 37 °C for extremely premature infants and 36.7 °C for very premature and late premature infants [BMR⁺20]. In this work, the term “body temperature” always refers to the body core temperature unless stated otherwise. In 1997, the WHO established definitions to quantify levels of severity of hypothermia (cold stress) and hyperthermia (high body temperature) for newborn infants. It should be noted that hypothermia is more common than hyperthermia nowadays, with a prevalence range of 32-85 % in hospitals and 11-92 % in homes [LBJH13]. Table 2.1 provides definitions for temperature ranges in newborns.

Tabelle 2.1: Definiton for temperature ranges in newborns, according to [Wor97].

Range	Definition
$< 32\text{ }^{\circ}\text{C}$	severe hypothermia
$32\text{-}36\text{ }^{\circ}\text{C}$	moderate hypothermia
$36\text{-}36.5\text{ }^{\circ}\text{C}$	mild hypothermia
$36.5\text{-}37.5\text{ }^{\circ}\text{C}$	normal range
$> 37.5\text{ }^{\circ}\text{C}$	hyperthermia

The usual range of body core temperature is defined as 36.5 °C – 37.5 °C. All temperatures above 37.5 °C are considered hyperthermia, while hypothermia is classified into three different stages. The range 36 °C – 36.4 °C indicates a mild hypothermia, which should raise concern.

If the body core temperature falls below 36 °C, this is called moderate hypothermia, which should require immediate rewarming of the infant. Moreo-

ver, below 32°C severe hypothermia is diagnosed, which can be very dangerous [Wor97], [KD14]. The different thermal stress levels of Figure 2.2 mostly follow the WHO standards. The only addition here is that Javed et al. introduced an additional level of severe hyperthermia for body core temperatures above 40°C [JM13].

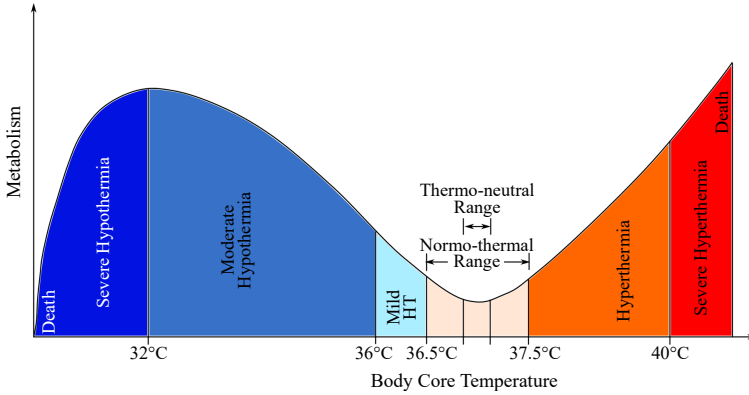


Abbildung 2.2: Metabolic activity depending on different thermal stress levels in newborns (adapted from [JM13]).

In the literature, different temperature boundaries are found, leading to a persistent disagreement on what is considered a normal body temperature in newborns. The lower boundary of the normal body temperature range suggested varies between 35.5°C – 36.5°C and the higher limit is proposed to be between 37°C – 38°C , but most studies and guides follow the WHO guidelines [LSC19], [Mic19], [MC⁺03]. However, the stages are defined and [Eur11] states that body temperatures below 35°C are fatal for extremely newborn infants. Regardless of the exact values, changes in the absolute central temperature or the central-peripheral temperature difference (cpTD) can indicate abnormal clinical states already [KHDSW09]. Asymmetrical temperature distributions of absolute temperature can also be interpreted. Because skin temperature distribution should be symmetrical across the body in humans, asymmetry can indicate a local abnormality in blood flow [KGR11].

In Figure 2.3, the *thermo-neutral range* and the *normo-thermal range* are shown. Both refer to the environmental temperature. The difference is that the thermo-neutral range is the range in which a human can maintain its body temperature without a significant increase in metabolism (36.8°C – 37.4°C for gestational age 28-29 weeks [KD14]), while the normo-thermal range describes

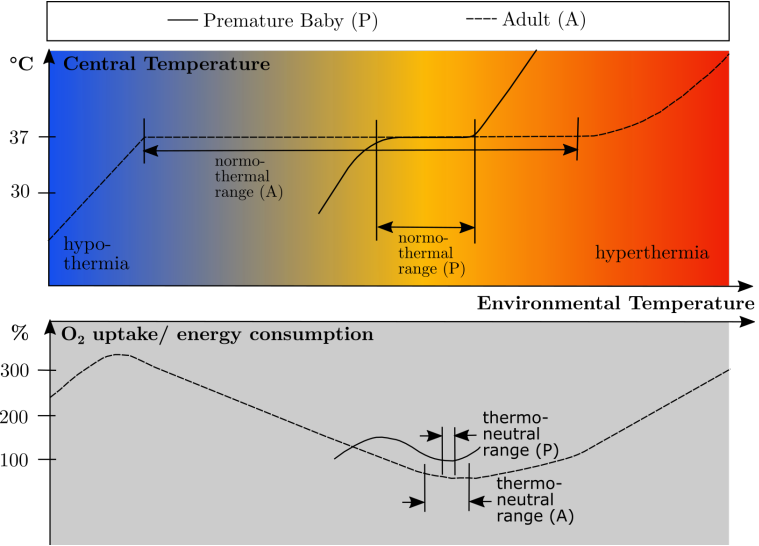


Abbildung 2.3: Metabolic reaction to changes in the environmental temperature and influence on the body temperature of adults and premature infants (adapted from [LP15]).

the span in which a human can maintain its optimal body temperature [LP15]. The thermo-neutral range is a subsection of the normo-thermal range. As seen in Figure 2.3, the latter is much wider for adults than for infants. Therefore, newborns and especially premature infants need a closely controlled ambient temperature in order to be able to maintain a body temperature inside the normo-thermal range as defined above. The necessary means for providing a controlled environment are described in chapter 2.1.3. After all, 37°C is a temperature that is optimal for cellular metabolism and has to be maintained as it implicates the minimal amount of oxygen and lowest energy consumption. Suppose the body temperature is increased or decreased. In that case, a higher metabolism can be detected, which is supposed to generate energy for heat production or reduction measures, as seen in Figure 2.2. Keeping the body temperature within the thermo-neutral range significantly reduces mortality and morbidity for low birth weight infants and infants born before 33 weeks of gestation. Additionally, it has beneficial effects on the control of breathing and heart rate and supports growth and sleep regulation [Bud07], [WM07], [LP15]. On the other hand, hyperthermia over a more extended period of time can

lead to seizures, coma, neurological damage and ultimately death. Untreated hypothermia may lead to apnoea, hypoglycaemia, hypoxia, coagulation defects, acute renal failure, necrotizing enterocolitis and finally death as well [WM07], [KDHDS⁺17]. The risk of falling into a state of hypothermia is higher outside hospitals. During this condition, mortality risk increases by 28% - 80% for every 1 °C decrease in body temperature, with the highest risk within the first 72 hours after birth. Hypothermia can happen in all climates, but it is more prevalent in winter than in summer [LSB⁺07], [LBJH13].

In order to avoid hypothermia, the newborn tries to counteract the thermal impact of its environment by balancing the heat losses, gains and production.

Premature infants lose heat through the skin and respiratory tract to the environment, based on four main environmental disturbance variables: the ambient temperature, the temperature of surrounding objects (T_{Surr}), the humidity and the air velocity (sometimes also the air pressure is counted). They can be grouped into the four central heat exchange mechanisms, which are radiation, evaporation, convection and conduction. The heat loss due to radiation is defined by

$$\dot{Q}_{Radiation} = A_{Skin,rad.} \cdot \sigma \cdot \epsilon_{skin} \cdot (T_{Surr}^4 - T_{Skin}^4), \quad (2.1)$$

being dependent on the radiative effective skin surface $A_{Skin,rad.}$, the Stefan-Boltzmann constant σ , the emission coefficient of the skin ϵ_{skin} , the skin temperature T_{Skin} and the temperature of the surrounding object T_{Surr} . As an example, a premature infant might gain heat from an overhead heater and could lose heat to a cold wall [LW16].

Convection describes the heat loss due to moving air and, therefore, depends on the ambient temperature. If the skin is warmer than the air, molecules rise from the skin and are dispersed. Convection can be defined by

$$\dot{Q}_{Convection} = A_{Skin,conv.} \cdot h_c \cdot (T_{Air} - T_{Skin}), \quad (2.2)$$

depending on the convectively effective skin area $A_{skin,conv.}$, the heat transfer coefficient h_c and the temperature difference between skin and air. The heat transfer coefficient increases with the ambient air velocity [LW16].

Conduction summarizes the heat flow due to skin contact with objects (e.g. incubator mattress) in a linear dependency:

$$\dot{Q}_{Conduction} = A_{Skin,cond.} \cdot K_c \cdot (T_{Mattress} - T_{Skin}), \quad (2.3)$$

being dependent on the heat conduction coefficient K_c between skin and incubator mattress and the contact area $A_{Skin,cond.}$. The heat conduction coefficient depends on the material of the mattress and the blood flow to the skin [LW16].

Evaporation happens when water is converted to gas and correlates with the trans-epidermal water losses and the ambient humidity.

Those four mechanisms are illustrated in Figure 2.4 [KD14], [LW16].

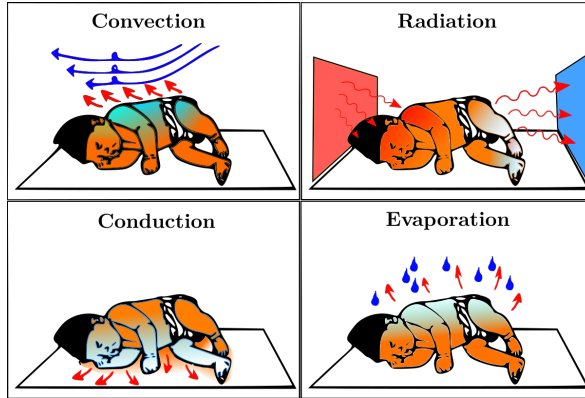


Abbildung 2.4: The four heat exchange mechanisms that can influence the body temperature (adapted from [LP15]).

Newborns can adjust their body temperature by using different thermal mechanisms. When they are too hot, they will speed up their metabolism and sweat more to cool down through evaporation and convection. However, most premature infants face the opposite problem of cold stress, due to low body temperature, which affects 38 % – 53 % of them [BMR⁺20], [LP15]. Cold stress can happen right after birth or during medical interventions. A body temperature below 36 °C will make the newborn narrow its blood vessels and burn brown adipose tissue (BAT) to generate heat without shivering. This increases the metabolism but also the oxygen demand. At the same time, the production of surfactant, a substance that helps the lungs function, decreases. This can lead to low oxygen and low blood sugar levels, which are harmful to growth [LP15]. Shivering, which involves muscle contractions, is not effective in young infants. The problem with using BAT is that it cannot grow outside the womb and will run out eventually [WM07]. Sometimes, doctors use controlled low body temperature (in the first three days of life) to protect the brain from damage, but this requires careful monitoring to avoid negative effects [Eur11].

Due to their impaired thermoregulation, premature infants have a poor ability to regulate their body temperature as they face several challenges. They lose heat faster because they have a larger surface area relative to their body mass (4.0-times the surface-per-weight of adults) than term-born infants

(2.7-times the surface-per-weight of adults) [LW16], [LSC19], [LBJH13]. They also have less subcutaneous fat and BAT, which provides insulation and heat production. Their skin is not fully developed, which causes more water loss, worse blood flow control and more evaporation [WM07]. Their heart, lung and muscle systems are also immature and affect their thermoregulation [LSC19]. These factors are especially severe in the first weeks and can cause a drop in body temperature of $0.2^{\circ}\text{C}/\text{min}$ to $1.0^{\circ}\text{C}/\text{min}$, which needs to be monitored closely for early detection and intervention [Bau08], [Wor97] [MVBDB08].

Therefore, premature and term-born infants, in general, are more likely to need external help than adults to maintain their body temperature and require constant thermal supervision [LP15], [WM07], [LBJH13],

2.1.3 Neonatal Incubators

One of the earliest devices for infant care was the couveuse, introduced in the 1880s. It was a wooden box with a platform for the infant to lie on. Underneath the platform, bottles filled with warm water created a warm and moist environment inside the box (see Figure 2.5).

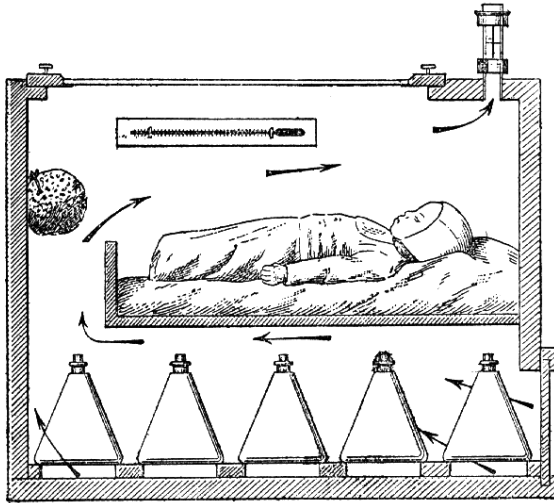


Abbildung 2.5: Operating principle of historical incubators using the example of Tarnier's Couveuse [Bud07].

Today, two therapeutic concepts for infant care are employed. In the closed

incubator, the infant is placed in an enclosed chamber (see Figure 2.6). The chamber can be controlled to have warm and humid air. The warm air should be distributed gently around the infant to prevent heat loss by convection. The humidity level should also be high to reduce evaporation losses through the infant's immature skin [Sed95].

In open care systems, the infant is only heated by a mattress and infrared heaters. These devices are not suitable for very premature infants and are only used when the infant has grown enough to regulate its own body temperature.



Abbildung 2.6: Closed Incubator [Wey23].

In incubators, the air temperature should be adjusted to maintain the thermoneutral zone for the premature infant. This requires monitoring the infant's body temperature. However, the body temperature is not uniform throughout the body [Koc16]. The central and peripheral body temperatures are measured to detect hypo- and hyperthermia. These two measurements can indicate whether the premature infant can keep the body temperature stable and whether

the premature infant is under thermal stress [LP15].

The difference between peripheral and body core temperature is slight in the thermoneutral zone. If the ambient temperature around the infant is too low, the peripheral temperature decreases faster than the body core temperature. The difference between the two temperatures increases and indicates cold stress. Likewise, heat stress causes the peripheral temperature to increase before the body core temperature [LP15]. A two-point measurement can, therefore, be used as an indicator of illness. If the premature infant has a fever, the body core temperature increases and the peripheral temperature decreases simultaneously [LP15]. The traditional method of measuring both temperatures uses temperature probes. This method has some drawbacks. First, the temperature probes are attached to the skin. Regularly changing the sensors can harm the immature, fragile skin, increasing the infection risk. Second, premature infants are often connected to many other devices by wires or tubes to ensure optimal care. Handling the infant with all these wires and tubes takes more time and causes more stress for the infant. Moreover, the large number of wires hinders the parents' bonding with their infant. However, bonding with the parents as early as possible is crucial for the infant's development [BBCL17].

Modern closed-loop incubators offer two different operating modes:

- **Air temperature control:** The air temperature is controlled to a set-point value. This is monitored by the nursing staff and continuously adjusted if necessary. Here, the patient acts as a disturbance variable on the control loop.
- **Skin servo control:** Here, the skin temperature, measured by a sensor, is regulated directly. The infant itself becomes part of the control loop. A cascaded control is often used, in which the air temperature is controlled in the inner control loop [TB99].

2.2 Infrared Thermography

The first one to propose the correlation between heat radiation and the emission of electromagnetic waves with wavelengths beyond those of red light was Sir William Herschel in 1800. In 1840 his son, J. F. W. Herschel, was the first to capture heat radiation on paper, calling it a *thermograph* [Bar63]. As this enabled the identification of living targets despite bad visual sight, the research and development regarding this topic mainly had been exclusive to the military until 1948, when the first medical thermogram, showing vascular patterns in limbs and breasts, was captured. Since then, this principle has been tested for detecting peripheral vascular diseases and related conditions such as breast

cancer and, more recently, has also to determine skin temperature distributions in premature infants [KGR11].

2.2.1 Infrared Radiation

The infrared band has its fixed place inside the electromagnetic spectrum, as depicted in Figure 2.7. It is commonly divided into near-infrared (NIR), short-wave infrared (SWIR), mid-wave infrared (MWIR), LWIR and far-infrared (FIR).

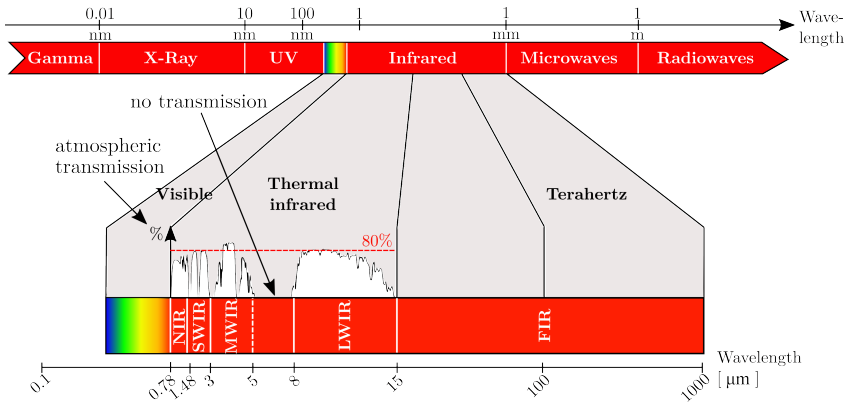


Abbildung 2.7: The infrared wavelength or frequency band within the electromagnetic spectrum including atmospheric transmission in the thermal infrared range (values based on [Inf23]).

A body can interact with electromagnetic waves via absorption or transmission, reflection and emission, as illustrated in Figure 2.8. While all those

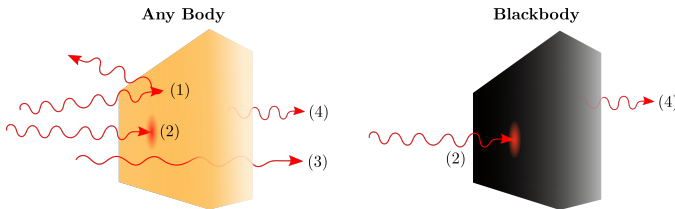


Abbildung 2.8: Possible interactions of radiation with tissue include reflection ρ (1), absorption (2) and transmission τ (3) and emission ϵ (4).

phenomena occur in typical bodies, there is the ideal concept of a black body, where only absorption and emission occur [Sch04].

The measurement principle of infrared thermography is based on the fact that objects with a temperature above absolute zero ($0\text{ K} = -273.15^\circ\text{C}$) emit electromagnetic waves, so-called *thermal radiation*, with a certain intensity Q and wavelength λ , with its spectrum varying based on the target's surface temperature T . The exact quantitative dependency has been formalized in *Planck's Law* using the natural constants for the speed of light (c), Planck's constant (h) and the Boltzmann's constant (k_b) as in eq. (2.4) [SW98]

$$E(\lambda, T) = \frac{2hc^2}{\lambda^5} \frac{1}{e^{\frac{hc}{\lambda k_b T}} - 1}. \quad (2.4)$$

The thermal radiation of human skin has its maxima above $8\text{ }\mu\text{m}$. Suppose a body has a temperature that leads to significant radiation intensity within the visual spectrum. In that case, the human eye can perceive this as colored (e.g., a glowing metal rod appears orange or the sun, a perfect black body with 5778 K appears white).

If the spectral intensity is integrated for each temperature individually, the following relationship between the total emitted radiation intensity Q_{total} and the temperature T can be found:

$$Q_{total}(T) = \sigma T^4. \quad (2.5)$$

Eq. (2.5) is called the *Stefan Boltzmann law* with the *Stefan-Boltzmann constant* being $\sigma = 5.67 \cdot 10^{-8} \text{ W m}^2 \text{ K}^{-4}$ [SW98]. Figure 2.9 depicts the spectral radiation intensity's dependency on the target's temperature.

It indicates that the wavelength of the intensity maximum is inversely proportional to the temperature of the emitting body, as can be found in *Wien's displacement law* (eq. (2.6)). Human skin at 32°C will have its heat radiation maximum at around $9.5\text{ }\mu\text{m}$.

$$\lambda_{max}(T) = \frac{2898\text{ }\mu\text{mK}}{T} \quad (2.6)$$

2.2.2 Infrared Sensors and Cameras

The contactless extraction of thermal information using IRT consists of two steps: the acquisition and the analysis. First, the uncompensated radiant thermal energy (heat) distribution emitted from a target surface has to be measured (apparent temperature) and second, the radiation intensity differences are calculated and corrected by further image processing.

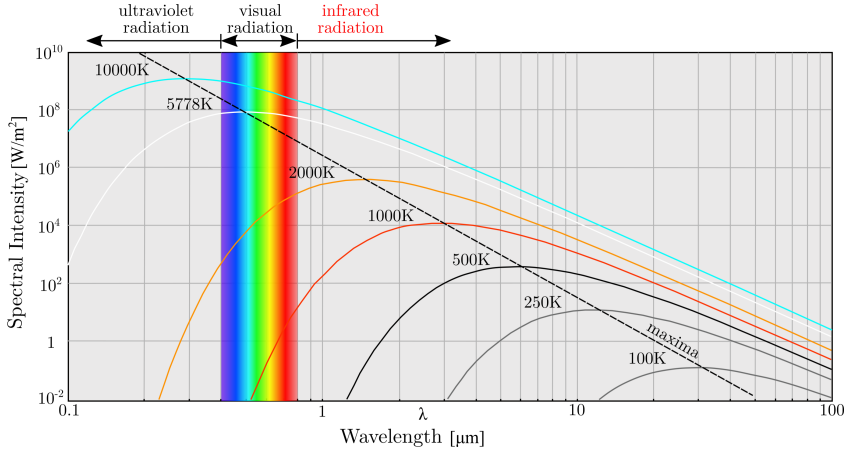


Abbildung 2.9: Dependency of spectral radiation intensity on the temperature of the target.

Nowadays, three different sensors are mainly used to measure the heat radiation. Thermopile sensors consist of a thin, blackened absorber that converts radiation into thermal energy and is positioned as a connecting piece on two different metals. Heating the thermocouple leads to an electromotive force (see Seebeck effect [VS86]) and a voltage can be measured. The sensitivity is comparatively low, but thermopile sensors are characterized by their robustness and low price [Rog12].

In a bolometer, the received radiation is also converted into heat. This method measures the temperature-dependent impedance change of the absorbing element. Microbolometers are often used because they are inexpensive and do not require active cooling [Rog12]. However, the lack of cooling also leads to fluctuations in the sensor temperature, which must be considered when interpreting the measured values.

In contrast to thermopile and bolometer sensors, which operate thermally, pyroelectric sensors measure the direct interaction of photons with the sensor. This type of sensor is, therefore, also called a photon detector. Incident photons interact directly with valence and quasi-free electrons in the detector crystal and the resulting voltage can be measured. Pyroelectric sensors are characterized by a high signal-to-noise ratio (SNR) and a short response time but require complex cryogenic cooling to suppress thermally induced charge carriers [Rog12].

Since pyroelectric sensors remain expensive, nowadays, microbolometers are often used to produce IRT cameras. They can be distinguished by field of view (FOV), spatial measurement resolution (in milliradians), detectable wavelength range, dynamic range, absolute temperature accuracy and price. Being able to capture a whole scene at once is also the main difference between thermal cameras and non-contact infrared thermometers [Adm21]. Starting from containing only one photosensitive cell, large, nitrogen-cooled, slow cameras with bad resolution and sensitivity, IRT cameras have become small, easily operable, uncooled, video-capable modules, which nowadays can even fit into smartphones.

2.2.3 Absolute Temperature Measurement

The ability to derive a target's temperature from its radiation highly depends on its emissivity [KGR11]. Thus, the measurement accuracy decreases with decreasing emissivity $\epsilon \in [0, 1]$ of the target surface, requiring a high emissivity of the target bodies for accurate absolute object temperature measurements. The best results can be achieved if the target surface belongs to a black body (ideal perfect emitter and absorber of thermal radiation at all wavelengths, $\epsilon = 1$) [KGR11] because then the fractions of reflected radiation (ρ) and transmitted radiation (τ) originating from the surrounding surfaces equal to zero according to

$$1 = \epsilon + \tau + \rho. \quad (2.7)$$

However, in reality, no body is a perfect black body [SW98] and the emissivity of objects varies with material combination, coating, roughness depth, viewing angle to surface normal, temperature and degree of polarization. Therefore, the radiation received by the infrared sensor also originates from other sources, such as the atmosphere or neighboring objects and needs to be compensated for [UVG⁺14]. Three different sources contribute to the total radiation W_{tot} received by an IRT camera:

1. Emission by the target object E_{obj}
2. Reflected emission by ambient objects/surroundings E_{surr}
3. Emission from the atmosphere E_{atm}

The total emission can be described by eq. (2.8):

$$W_{tot} = E_{obj} + E_{surr} + E_{atm} \quad (2.8)$$

Figure 2.10 depicts a schematic representation of the composition of the total emission.

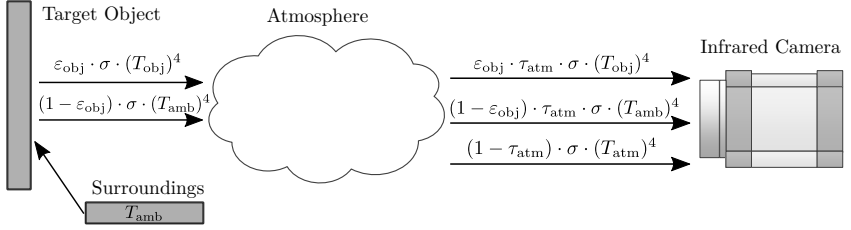


Abbildung 2.10: Total Radiation received by an infrared camera adapted from [UVG⁺14].

The object itself is the first source of infrared radiation. However, not all of the radiation emitted by the object reaches the camera. Due to the transmission coefficient of the atmosphere (τ_{atm}), some of this emitted energy is absorbed by the atmosphere before it reaches the camera. The emission from the target object can be expressed using eq. (2.9)

$$E_{\text{obj}} = \varepsilon_{\text{obj}} \cdot \tau_{\text{atm}} \cdot \sigma \cdot (T_{\text{obj}})^4, \quad (2.9)$$

with ε_{obj} being the emissivity of the target object, σ the Stefan-Boltzmann constant and T_{obj} the target's surface temperature.

Unlike blackbodies, greybodies have a reflectivity greater than zero. This means that they reflect some of the infrared radiation emitted by their surroundings. However, some of this reflected radiation is also absorbed by the atmosphere. This is the second source detected by the camera and can be expressed as eq. (2.10)

$$E_{\text{surr}} = \rho_{\text{obj}} \cdot \tau_{\text{atm}} \cdot \sigma \cdot (T_{\text{surr}})^4 = (1 - \varepsilon_{\text{obj}}) \cdot \tau_{\text{atm}} \cdot \sigma \cdot (T_{\text{surr}})^4 \quad (2.10)$$

where $\rho_{\text{obj}} = (1 - \varepsilon_{\text{obj}})$ describes the reflectivity of the target object and T_{surr} the temperature of the ambient sources.

The last source to reach the camera is the infrared radiation emitted by the atmosphere. We can express this using eq. (2.11), where $(1 - \tau_{\text{atm}})$ represents the emittance of the atmosphere

$$E_{\text{atm}} = \varepsilon_{\text{atm}} \cdot \sigma \cdot (T_{\text{atm}})^4 = (1 - \tau_{\text{atm}}) \cdot \sigma \cdot (T_{\text{atm}})^4 \quad (2.11)$$

with T_{atm} and ε_{atm} representing the temperature and emissivity of the atmosphere.

Substituting eqn. (2.9)-(2.11) provides the total emission received by the camera.

$$W_{\text{tot}} = \varepsilon_{\text{obj}} \cdot \tau_{\text{atm}} \cdot \sigma \cdot (T_{\text{obj}})^4 + (1 - \varepsilon_{\text{obj}}) \cdot \tau_{\text{atm}} \cdot \sigma \cdot (T_{\text{surr}})^4 + (1 - \tau_{\text{atm}}) \cdot \sigma \cdot (T_{\text{atm}})^4 \quad (2.12)$$

Solving the equation for T_{obj} yields the temperature of the target object:

$$T_{\text{obj}} = \sqrt[4]{\frac{W_{\text{tot}} - (1 - \varepsilon_{\text{obj}}) \cdot \tau_{\text{atm}} \cdot (T_{\text{surr}})^4 - (1 - \tau_{\text{atm}}) \cdot (T_{\text{atm}})^4}{\varepsilon_{\text{obj}} \cdot \tau_{\text{atm}} \cdot \sigma}} \quad (2.13)$$

In order to solve the equation, specific parameters are required. These include the emissivity of the object ε_{obj} , the reflected temperature T_{surr} , the transmittance of the atmosphere τ_{atm} and the temperature of the atmosphere T_{atm} . The transmittance of the atmosphere is typically estimated from the object's distance from the camera and the relative humidity. This value is usually very close to one. The atmospheric temperature is measured using a standard thermometer. However, as the emissivity $\varepsilon_{\text{atm}} = 1 - \tau_{\text{atm}}$ of the atmosphere is close to zero, this parameter has minimal effect on the temperature measurement. On the contrary, the emissivity of the object and the reflected temperature significantly affect accurate temperature measurements and must be measured carefully [UVG⁺14, Hol00].

In the case of a newborn child, the emissivity (ε_{obj} for eq. (2.13)) can be assumed as 0.98 [LSC19], which is close to a black body, making it reasonable for surface temperature measurement. Additionally, the radiation maxima of human skin ($-10^\circ\text{C} - 50^\circ\text{C}$) occur within the wavelength range $9\mu\text{m} - 11\mu\text{m}$. This aligns with the capabilities of IRT cameras operating in the MWIR or LWIR regions, where the wavelength spans from $2\mu\text{m} - 14\mu\text{m}$ (see figure 2.7).

2.3 Deep Learning

In traditional computer algorithms, a human programmer unambiguously specifies the steps a machine has to take to solve a given problem. However, not every problem can be solved using such clearly definable steps and rules. Either because these rules do not exist or because the implementation is too complex to be identifiable or implementable with reasonable effort at the current state of science [Bie18].

Machine learning is an alternative approach to such problems in which the computer approximates rules from the data itself. Instead of providing rules and data, the human programmer provides data and the desired output in this method. Based on this, the program derives and learns rules, which can

then be applied to information the program has not seen before (see figure 2.11) [Cho18].

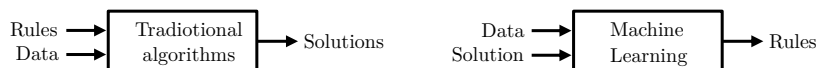


Abbildung 2.11: Difference between traditional algorithms and machine learning (based on [Cho18]).

Deep learning is a subset of machine learning that uses artificial neural networks, inspired by how the human brain works, to learn from complex and large data sets.

2.3.1 Artificial Neural Networks

Today’s most common way to implement deep learning is an artificial neural network [Cho18]. A basic neural network consists of subsequent connected layers that are, in turn, comprised of individual neurons. A neuron is a mathematical building block that processes pieces of input data and transforms them into an output signal. First, the individual neuron’s inputs are weighted and then merged together using the input function (mostly a sum). This merged input is then further processed by an activation function, which calculates the neuron’s activation. This activation, in turn, functions as input for the following neuron [Bie18, GT19].

The most commonly used activation functions are visualized in Figure 2.12

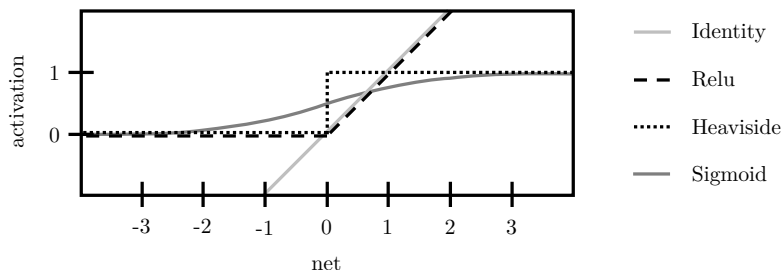


Abbildung 2.12: Visualization of the most common activation functions, based on [Bie18, Cho18]).

2.3.2 Convolutional Neural Networks

For image processing, however, using networks with only fully connected (dense) layers consisting of simple neurons poses challenges. On one hand, the number of trainable parameters and, thus, the time and cost of training the networks are high. On the other hand, this network architecture and the transformations happening in its layer are not specifically designed to process spatial patterns in 2D images. Instead, a convolutional neural network (CNN) is the most commonly used network architecture for computer vision tasks. A simple CNN consists of several convolutional and pooling layers, which are arranged alternately. This structure is traditionally completed by several fully connected layers which perform the final classification [Cho18,SSM⁺17].

A convolutional layer individually transforms small image sections rather than the entire image simultaneously. The process is presented in simplified form in Figure 2.13.

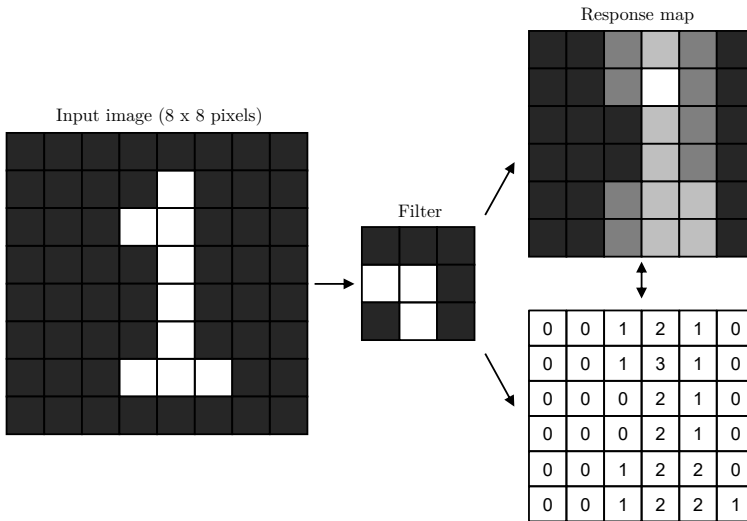


Abbildung 2.13: Simplified example of a convolution operation (based on [Cho18,Suk17]).

A filter of size $m \times m$ (usually $m = 3$ pixels or $m = 5$ pixels) is slid over the image and the respective image patch is transformed at each location. Put together, the individual responses create a response map. This is repeated with

several different filters in every convolutional layer. The response map of one layer put together, in turn, yields a feature map that forms the output of a convolutional layer [Cho18, SSM⁺17, Suk17]. In a pooling layer, these feature maps are then reduced in size. This is done by taking each response map and combining a region of $d \cdot d$ values into one digit. For this, the maximum (max-pooling) or the mean (average-pooling) of the d^2 values is mostly chosen [SSM⁺17].

In both pooling and convolutional layers, the stride parameter defines how far the $d \times d$ window or the filter is moved between each calculation step [SSM⁺17, Suk17]. The concept of max-pooling is illustrated in Figure 2.14 for two different strides. In Figure 2.13 a stride of one/two and $d = 2$ is used.

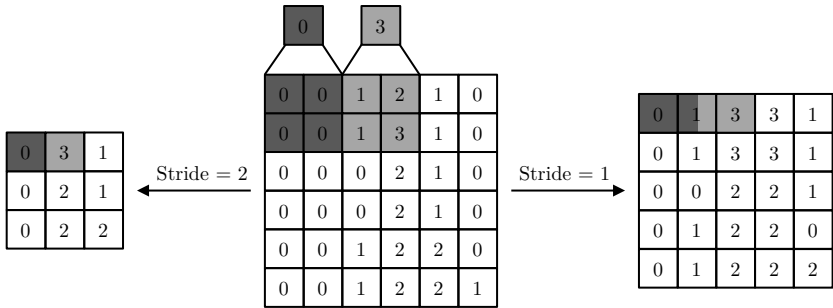


Abbildung 2.14: Illustration of a pooling operation for two different strides (based on [Suk17]).

In comparison to a network made solely of dense layers, a CNN can learn local rather than just global patterns in a picture. This allows for a reduction of the trainable parameters as individual local patterns are invariant to their global position in a picture (translational invariance). Furthermore, the combination of pooling and convolutional layer enables the network to learn hierarchies of patterns and thus detect more intricate structures [Cho18, SSM⁺17, Suk17].

2.3.3 Semantic Segmentation

Computer vision describes the concept of analyzing visual data such as photos or videos using a computer [Sze11]. Typical tasks are the classification of images into different categories or the detection of objects in images [SSM⁺17]. Semantic segmentation is another typical computer vision problem, which refers to defining a set of classes and assigning every pixel in an image to one of these classes [HWW⁺19]. Human Parsing, in turn, is an application of semantic

segmentation in which these class labels comprise human body parts and/or clothing items [LWL⁺20, RLH⁺19].

A further distinction can be made between instance segmentation and global human parsing. While the former distinguishes between different instances of the same object (e.g. different people), global human parsing differentiates between body parts but not humans. Two pixels showing a head are consequently assigned to the same class, regardless of whether they display the same or a different person. Multiple human parsing in turn combines these two types of segmentations by distinguishing both body parts and instances, see Figure 2.15. [LXWY20]

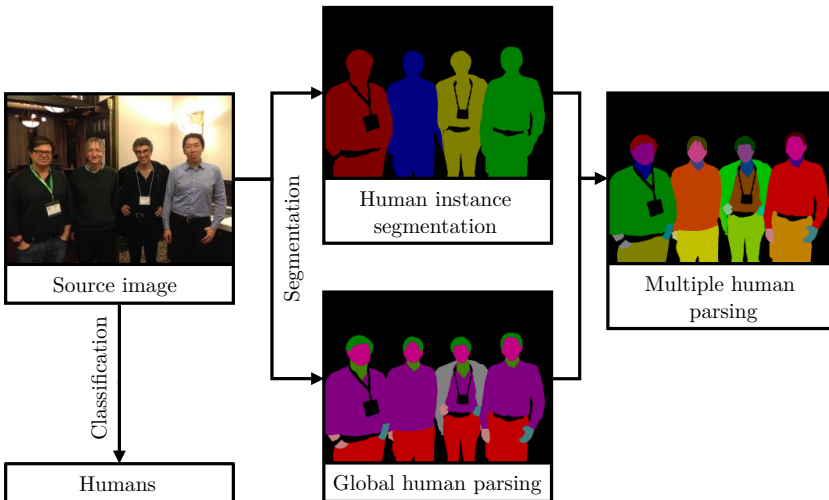


Abbildung 2.15: Differences between image classification and segmentation and different kinds of human parsing problems (based on [LXWY20]).

In semantic segmentation tasks, the localization of objects in an image is required [RFB15]. After downsampling the image through several convolutional and pooling layers (encoder structure), the image, therefore, needs to be reconstructed by a decoder [ZZP⁺18]. Such upsampling can be done using an operation referred to as backward convolution or deconvolution. One method for upsampling is interpolating the outputs from the nearest inputs. Hence, this process is also described as convolution with a stride between zero and one [LSD15].

2.3.4 Training and Overfitting

Various training and learning mechanisms exist, each suited for different tasks regarding neural networks [Suk17]. Supervised learning is generally used for the task of image segmentation. Here, the computer is given input data and the desired outputs (targets), is typically provided by human operators. From this, the program learns functional transformations for the input data to map them to the targets.

Suppose all the weights of a neural network are randomly chosen numbers. In that case, any input data is transformed arbitrarily and the network yields an output that does not equal the desired solution. Therefore, the first step in a training cycle compares the network's current output with the desired result to a given input. This is done using the loss function(L), which calculates how well the network approximates the solution. Based on the computed loss value, an optimizer algorithm then updates the weights of the network in a way that reduces the loss. However, the set of parameters that minimizes the loss function usually cannot be found analytically due to the high number of weights in conventional neural networks. Therefore, the weights are updated iteratively using a gradient descent approach. First, the gradient of the loss function is evaluated at the given set of weights. As stated above, a neural network consists of interconnected layers, which can be summarized as a chain of subsequent transformations. Therefore, the gradient is calculated by applying the chain rule multiple times to propagate the error from the output layer back through the individual layers step by step (backpropagation algorithm). In a network with M layers, the gradient of the loss function (L) as a function of the weights in layer m ($\mathbf{W}^m$) can hence be calculated using eq. (2.14).

$$\frac{\partial L}{\partial \mathbf{W}^{(m)}} = \frac{\partial L}{\partial \mathbf{a}^{(M)}} \frac{\partial \mathbf{a}^{(M)}}{\partial \mathbf{a}^{(M-1)}} \cdots \frac{\partial \mathbf{a}^{(m+2)}}{\partial \mathbf{a}^{(m+1)}} \frac{\partial \mathbf{a}^{(m+1)}}{\partial \mathbf{a}^{(m)}} \frac{\partial \mathbf{a}^{(m)}}{\partial \mathbf{x}^{(m)}} \frac{\partial \mathbf{x}^{(m)}}{\partial \mathbf{W}^{(m)}} \quad (2.14)$$

The parameters are subsequently updated in the negative direction of the gradient, slightly reducing the loss (see eq. (2.15)). How much the weights are changed per iteration step (τ) is determined by the learning rate (lr).

$$\mathbf{W}^{(\tau+1)} = \mathbf{W}^{(\tau)} - lr \nabla L(\mathbf{W}^{(\tau)}) = \mathbf{W}^{(\tau)} - lr \left[\frac{\partial L}{\partial \mathbf{W}^{(1)}} \cdots \frac{\partial L}{\partial \mathbf{W}^{(m)}} \cdots \frac{\partial L}{\partial \mathbf{W}^{(M)}} \right] \quad (2.15)$$

Finally, the new output of the network to a given input is calculated and the described training cycle starts again. If repeated sufficiently often, this cycle optimizes the weights and minimizes the loss. The general process for training a neural network is exemplified in Figure 2.16 [Cho18, Suk17].

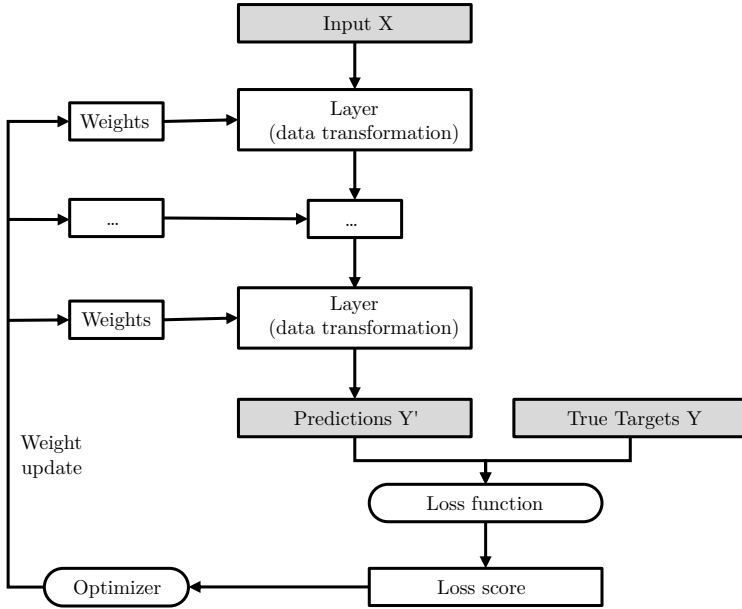


Abbildung 2.16: Basic training cycle for a neural network (based on [Cho18]).

The training process aims to adjust the weights of the network to find expedient transformations for the input data. However, the training does not aim to find the best combination for the given training samples but rather to enable the network to detect general patterns that can be transferred to new data as well. Nevertheless, the weights are updated solely based on knowledge from the given training samples. This leads to a conflict between optimizing the network's performance on the training data and generalizing learned patterns. If the model is overfitting, the network improves its training performance by learning less important and sample-specific patterns, which may deteriorate its generalization and thus its performance on new data (see Figure 2.17) [Smi18], [Cho18].

Increasing the size of the training dataset is the preferred way to alleviate this problem, as this also increases the amount of available information to learn from during training. Since this is not always possible with reasonable effort, several other methods to mitigate overfitting exist. A selection of commonly used strategies (i.e. data augmentation, dropout, hyperparameter optimization, batch normalization and transfer learning) is described in the following [Cho18].

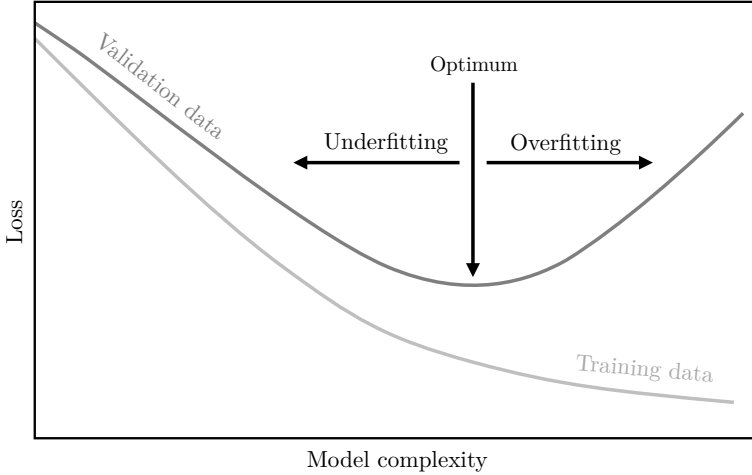


Abbildung 2.17: Illustration of overfitting and underfitting (based on [Smi18]).

Data augmentation is a method to artificially enlarge the training dataset by modifying the existing data. By e.g. shifting, cropping, rotating, or changing the colors of the training images, more general features can be learned from the same set of pictures. Furthermore, dropout aims to force the network not to rely on individual neurons or features to improve generalization. This is achieved by randomly selecting a specified percentage of the neurons or features and setting their activation to zero during a training cycle [PW17,SK19].

Hyperparameters are all the parameters of a model that are not the trainable weights (e.g. model architecture, learning rate, number of layers, etc.). If the number of trainable weights is too high, for example, the network has free capacity to learn irrelevant patterns. If the number of weights is too low on the other hand, the network may not be able to learn all the essential features. Therefore, the hyperparameter of the network must be optimized for the task and data at hand. Besides randomly trying and comparing different parameters manually (random search), several methods for automizing the hyperparameter optimization have been developed. However, most of these automized methods are inefficient, costly, or not reliable in the current state of research [Cho18, HKV19, IAFS17].

Moreover, batch normalization pursues the goal of limiting the differences between activations inside the individual layers of a network by normalizing layer outputs. Usually applied between convolutional and pooling layers or

between dense layers, this method can accelerate convergence and allow for the use of a higher learning rate [Cho18, IS15, SK19].

Lastly, transfer learning uses existing large datasets created for similar tasks to pre-train a network for the problem at hand. This strategy is often used in computer vision as basic spatial features are not dependent on the exact task or content of the images and can, therefore, be learned well from large but non-task-specific datasets. The pre-trained network can subsequently be adapted with small but specific datasets to solve the given problem. Feature extraction refers to reusing the entire pre-trained model except for the last layer(s) so that only this classifier is trained on task-specific training data. This is reasonable, for example, if the classification task for which the pre-training dataset was created is similar to the given task but the number of classes to be distinguished differs. During fine-tuning, in contrast, more layers, in addition to the classifier, are trained on the task-specific data. Often, the last layers of a CNN learn task-specific representations, while the first layers detect basic spatial patterns (e.g. edges). Therefore, the more the tasks of the training data differ, the more layers may need to be fine-tuned [Cho18, PW17, SK19, TSG⁺16].

2.4 Control Theory

Figure 2.18 represents a standard control in the Laplace domain. It consists of the transfer functions of the controller $G_C(s)$ and the controlled system $G_S(s)$. Together, both form the transfer function of the open control loop $G_0(s) = G_C(s) \cdot G_S(s)$. The reference variable $W(s)$ specifies a setpoint value that the controlled variable (actual value) $Y(s)$ should reach. The controlled variable is fed back negatively to the input via summation, which results in a control error $E(s)$ in the event of deviations between $Y(s)$ and $W(s)$. This is the input of the controller. Depending on the control error, the controller calculates the manipulated variable $U(s)$, which is the input of the controlled system.

In addition, an additive disturbance at the output $Z(s)$ is modeled here, which influences the controlled variable. The modeling of disturbances can be significantly more complex, depending on the application [Unb08].

If $W(s)$ is constant, the control is a fixed value control. Ideally, $Y(s)$ should correspond to $W(s)$. The task of the controller is to suppress disturbance variables as well as possible [Unb08].

In order to design a controller $G_C(s)$, a mathematical model of the controlled system $G_S(s)$ must be approximated. This can be done either by mathematical modeling or by analyzing suitable measurements of the actual process [Lju99]. The subsection System Identification explains how transfer functions may be

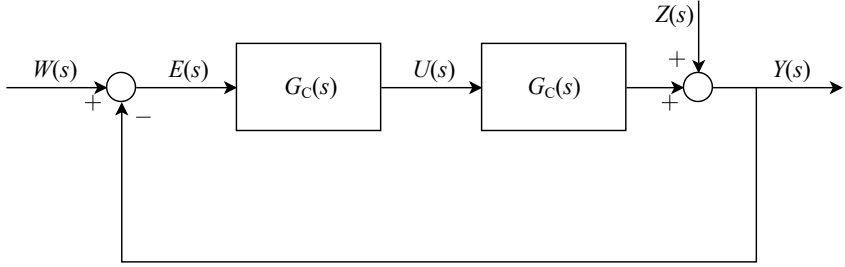


Abbildung 2.18: Standard control loop.

approximated from measurement data. As the identified controlled systems are often ambiguous, the uncertainty subsection explains how these are modeled and how they can be considered when designing a controller. Finally, the last subsection introduces the idea of a hardware-in-the-loop test bench.

2.4.1 System Identification

In order to identify a system, suitable data must be collected. The input of the system (i.e. the controlled system) $u(t)$ and the output $y(t)$ must be recorded. The aim is to obtain a suitable mathematical model $G_S = \frac{Y}{U}$.

Simple measurement data for system identification are, for example, step functions that are recorded in the value range of interest for later control [Lju99].

The first step is to define a suitable type of system. One category of systems is a linear time-invariant (LTI) system. It consists of a static gain K_g , pole points with time constants $T_{p1}, \dots, T_{pn}$, zero points with time constants $T_{z1}, \dots, T_{zn}$, a dead time T_t and optionally an integrator. A simple PT_1T_t element can be described by:

$$G_S(s) = \frac{K_g \cdot e^{-T_t s}}{1 + T_p s} \quad (2.16)$$

LTI systems are parametric systems, as their structure is already defined and only the parameters (i.e. time constants and gain) need to be estimated. The output $\hat{y}(t, \theta)$ of the estimated system $\hat{G}_S$ should correspond as closely as possible to the measured output $y(t)$. Here, θ stands for all parameters that are to be estimated [Lju99].

If input and output data ($u_{\text{meas}}(t)$ and $y_{\text{meas}}(t)$) are available, the difference $\epsilon(t, \theta) = y_{\text{meas}}(t) - \hat{y}(t, \theta)$ must be minimized by varying θ , where $\hat{y}(t, \theta) = \mathcal{L}^{-1}\{\hat{G}_S(s)U_{\text{meas}}(s)\}$ applies. For this purpose, a cost function $V(\theta)$ to be mi-

nimized is defined:

$$V(\theta) = \frac{1}{N} \cdot \sum_{t=1}^N \epsilon(t, \theta)^2 \quad (2.17)$$

This cost function is minimized using numerical methods. This results in parameters that minimize this cost function and represent an optimal approximation under these conditions. It is important to note that the approximated system only ever represents an approximation of reality and that the selected model must be suitable so that this approximation is sufficient [Lju99].

2.4.2 Uncertainty Modeling

In real-world systems, it is often impossible to know all system parameters with absolute certainty. These uncertainties can arise from various sources, such as:

- Measurement errors: Sensors and measurement devices are inherently subject to noise and inaccuracies.
- Parameter variations: System parameters may change over time due to environmental factors, wear and tear, or other dynamic processes.
- Model simplifications: Mathematical models used to represent complex systems are often simplified representations of reality, leading to inherent uncertainties.

Ignoring these uncertainties during control design can lead to controllers that perform poorly or even cause instability when implemented in real-world applications. Therefore, it is crucial to model and account for uncertainties during the control design process. Parametric uncertainty is a way of modeling uncertainties. For this purpose, parameters $x = x_0 + \delta_x$ are defined, which consist of a *nominal value* x_0 and a *variation* δ_x . Here, δ_x can assume a fixed value range and thus ensures that x does not have a fixed value but is uncertain. Uncertain parameters can then be used to represent the uncertainty of parametric models [GPK13].

As an example, a PT_1 element with the transfer function $G(s) = \frac{K}{1+T_s}$ is considered. Here, the static gain is $K_g = K_0 + \delta_K$, with $K_0 = 1$ and $\delta_K \in [-0.1, +0.1]$. As for the time constant: $T_p = T_0 + \delta_T$, with $T_0 = 10 \text{ sec}$ and $\delta_T \in [-2 \text{ sec}, +2 \text{ sec}]$. The time constant T_p can, therefore, assume all values in the interval $[8 \text{ sec}, 12 \text{ sec}]$ and the gain K_g all values in the interval $[0.9, 1.1]$.

Figure 2.19 depicts ten step responses of the uncertain system. The figure illustrates the modeled uncertainty.

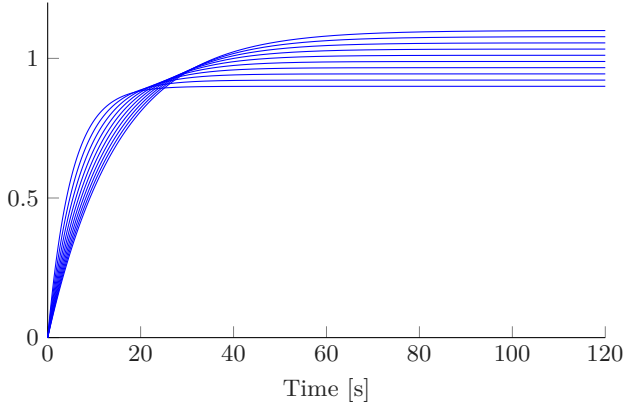


Abbildung 2.19: Step responses of an uncertain system $G(s) = \frac{K_g}{1+T_p s}$.

When considering stability, *nominal stability* means that the nominal system is stable. *Robust stability* means that each system represented by the uncertainty is stable [ZD].

When designing a controller for an uncertain model, it is therefore essential to strive for robust stability. This can be determined numerically. It also dictates the extent to which uncertainty can increase without causing system destabilization [GPK13].

2.4.3 Hardware-in-the-Loop

Hardware-in-the-Loop is a powerful testing methodology used to validate and refine the performance of complex systems, particularly those involving embedded systems and real-time control. It bridges the gap between purely software-based simulations and real-world testing by integrating a physical component (the "hardware") with a virtual environment (the "loop") [ISS99, Bac05]. In a HIL setup, the hardware under test (e.g., an electronic control unit or a sensor) is connected to a simulator that emulates the real-world environment in which the hardware would normally operate, as depicted in Figure 2.20 .

The simulator generates stimuli and inputs for the hardware, mimicking real-world conditions and scenarios. The hardware then processes these inputs and produces outputs, which are fed back into the simulator. This creates a closed loop where the hardware interacts with a virtual representation of its environment in real time [NBAR04].

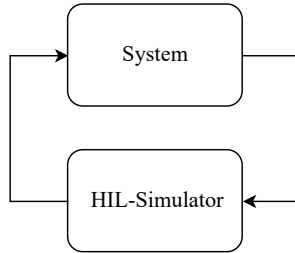


Abbildung 2.20: Hardware-in-the-Loop.

HiL testing offers several advantages over traditional testing methods:

- **Reduced risk:** HiL allows for testing potentially dangerous scenarios in a safe and controlled environment before deploying the hardware in the real world. This is particularly valuable for systems where failures could have severe consequences.
- **Increased efficiency:** HiL testing can be significantly faster and more cost-effective than real-world testing, as it eliminates the need for physical prototypes and extensive field testing.
- **Improved test coverage:** HiL allows for comprehensive testing of various scenarios and edge cases that might be difficult or impossible to replicate in real-world testing.
- **Enhanced debugging:** HiL setups facilitate efficient debugging by allowing developers to observe the hardware's behavior in real-time and analyze its response to specific stimuli.

HiL testing is widely used in various industries, including:

- **Automotive:** Testing engine control units, anti-lock braking systems, and other critical components.
- **Aerospace:** Validating flight control systems, avionics, and other safety-critical systems.
- **Medical devices:** Testing the functionality and safety of medical equipment and control systems.
- **Robotics:** Evaluating the performance of robot control algorithms and hardware components.

3 Body Part Segmentation of Premature Infants using a Multi-Modal Camera System

Monitoring the body temperature of premature infants is essential for their optimal management. Thermography provides a wireless alternative to traditional wired methods. Due to infant movement, automatic segmentation of body regions is critical. This chapter presents a novel deep learning algorithm for infant body region segmentation. Four neural networks based on a U-net architecture were compared. The first two networks used a single imaging modality (visible light or thermography), while the last two networks fused features from both imaging modalities. The dataset consisted of 600 visible light and 600 thermography images of 20 infants. The images were manually annotated. Transfer learning and data augmentation improved the segmentation performance. The fusion models performed best (mean Intersection over Union of 0.87), followed by the RGB model (mean Intersection over Union of 0.85). The thermography model, although slightly less accurate (mean Intersection over Union of 0.75), still provided good results. In conclusion, multi-modal neural networks offer a novel approach to body segmentation of premature infants using feature fusion and cross-modality transfer learning with limited data. Parts of this chapter have already been published in peer-reviewed journals and conferences [VBL⁺23], [VKLL24].

3.1 Motivation and Overview

Current practice for monitoring the temperature of premature infants relies on wired thermistors glued to the skin with strong adhesives. However, this approach can have adverse effects, promoting bacterial growth, skin damage and irritation. In addition, removal of these adhesives can cause skin tears, further compromising the infant's delicate skin barrier function [BDL⁺01]. Beyond the direct skin contact problems, the cables themselves are a significant burden. They interfere with transport, and achieving optimal positioning within the incubator becomes a challenge due to cable clutter. Notably, studies suggest that the presence of these cables can even create a psychological barrier for parents interacting with their child [BBCL17]. These limitations highlight the urgent need for a wireless and non-contact alternative for temperature monitoring.

One promising solution is infrared thermography. This technology measures the infrared radiation emitted by the skin to determine temperature. However, to accurately extract the infant's temperature distribution, automatic segmentation of the skin within the region of interest (ROI) is critical, as the infant's movements can cause these regions to shift.

The field of body part segmentation (also known as human parsing) has ma-

de significant progress in recent years, with benchmarks achieving impressive improvements [RLH⁺19, YSWJ19, ZLC⁺18]. However, research on body part segmentation specifically in premature infants remains a gap in the field. Their body proportions differ significantly from those of adults, posing a unique challenge. While studies by Chaichulee et al. [CVJ⁺17], Dossso et al. [DBN⁺18] and Villaroel et al. [VCJ⁺19] segmented infant skin, they did not differentiate individual body parts. Zhang et al. [ZXH19] attempted to segment body parts in infant videos to identify potential neurological problems based on movement patterns. However, their limited dataset size hindered the ability of the neural network to generalize to unseen videos.

Hoog Antink et al. [AFP⁺20] improved the task of segmenting the body parts of premature infants using a neural network. Their approach used a dataset of 643 manually annotated images captured in both RGB and near-infrared (NIR) spectra. The aim was to enable non-contact heart rate monitoring. Their results were particularly promising for head segmentation, with an average Intersection over Union (IoU) of 0.82. However, pre-training on extensive datasets of adult human parsing proved critical to this success. Training on 81 NIR images alone resulted in significantly less reliable segmentation, highlighting the limitations of smaller, specialized datasets.

A novel approach to body part segmentation with LWIR images of premature infants was presented by Asano et al. [AHH⁺22]. Their method combined a U-network architecture with a Generative Adversarial Network and a Self-Attention module. Notably, they did not rely on pre-training their neural network, but instead used a dataset specifically containing 400 images of premature infants.

All of the above studies used only single modalities to segment the body. However, the fusion of different modalities has demonstrated remarkable success in different research areas. For example, Farahnakian et al. [FH20] investigated early, hybrid and late fusion strategies for maritime vessel detection and compared them with networks using only RGB or LWIR images. All fusion networks outperformed single-modality approaches, with RGB networks achieving higher accuracy than those using only LWIR.

Early fusion techniques combine information from two images at the pixel level before further processing. This can be achieved by concatenating an RGB and LWIR image into a four-channel image that is fed into the neural network. Conversely, late fusion combines information at the decision or classification level. Separate neural networks process RGB and LWIR images and the outputs are then concatenated or fused [ZCVZ15].

The effectiveness of fusion strategies has been further validated by Sun et al. [SZL19]. These authors used a feature fusion approach with two encoders and one decoder for semantic segmentation tasks, fusing RGB and LWIR images

of urban scenes. Notably, no existing research directly compares the use of RGB images, LWIR images, or a combined multimodal approach for body part segmentation in premature infants.

This chapter presents a multimodal approach to infant body part segmentation. Data recently acquired in a clinical trial by both RGB and LWIR cameras was used. Based on this dataset, different U-net-based approaches for segmentation were compared. The first approach used a single modality, feeding information from either the RGB or LWIR camera into a standard encoder-decoder architecture. In addition, different hybrid fusion approaches were examined that combine information from both modalities to improve performance.

In this chapter, Section 3.2.1 will describe the available data and the specific datasets are chosen. Furthermore, in Section 3.2.2 the neural network architectures and fusion approaches will be outlined. Subsequently, in Sections 3.2.3 and 3.2.4 preprocessing, training and validation strategies will be explained. Section 3.3 then will present the results that the chosen hyperparameters and data augmentation, transfer learning and fusion strategies achieve on the segmentation task. Finally, the chosen models will be validated on unseen test data and in Section 3.4 the results will be discussed.

3.2 Materials and Methods

3.2.1 Dataset

The dataset utilized in this chapter was collected by Lyra et al. [LRH⁺22] in the neonatal intensive care unit of Saveetha Medical College & Hospital, Chennai, India. The study was approved by the institutional ethics committee of Saveetha University (SMC/IEC/2018/03/067). It included 19 stable patients with gestational age at birth ranging from 29 to 40 weeks. As one patient was recorded twice, there are a total of 20 different recordings. The infants ranged in weight from 1500 g to 3010 g, and in age from 37 hours to 56 days after birth.

RGB images were captured using a Grasshopper 3 GS3-U3-23S6C-C camera (FLIR, USA) with a frame rate of 60 Hz and a resolution of 1920 x 1200. Simultaneously, IR images were acquired with a VarioCAM HD head 820 S camera (InfraTec, Germany) at 10 Hz and a resolution of 1024 x 768. The spectral range of the thermal detector is from 7.5 to 14.0 μm , which is in the LWIR region. The premature infants were imaged in an open incubator.

From these recordings, 30 images per recording (600 in total) were manually selected according to three specific criteria

1. Careful consideration was given to selecting images that were not taken

immediately after each other so that images from the same shot were not redundant.

2. Images were selected to include all postures, movements and other activities, such as interaction with the clinician.
3. Since the cameras filmed at different frequencies, the chosen RGB and LWIR images were ensured to depict the same point in time.

The RGB images were labeled using the MATLAB Image Labeler application (The MathWorks, Inc.). The approach in this chapter differs from previous research [AFP⁺20, OVB⁺16], where seven classes ("background", "head", "torso", "left arm", "right arm", "left leg", "right leg") were commonly used. As the central-peripheral temperature difference is typically measured with only two probes, it is not common to distinguish between left and right body parts. Therefore, a simpler labelling scheme with five classes was defined: "background", "head", "torso", "arm/leg". In addition, only visible skin was labeled as part of the body, while clothing was considered background (see Figure 3.2a)). This dataset is referred to as the Chennai dataset.

However, the Chennai dataset has limitations. Because it only includes images of infants from a specific region of India, it may not include the full range of human skin tones. This could potentially bias deep learning models trained on this data. In addition, the radiant heaters in the incubators were on for seven of the infants, but off for the others. This inconsistency causes variations in the contrast between body parts and the background in the LWIR images, see Figure 3.1.

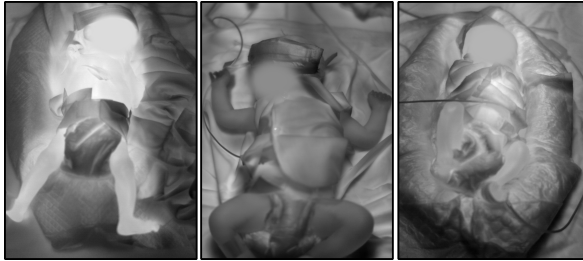


Abbildung 3.1: LWIR recordings with different contrasts due to the radiant heater.

When the heaters were on, the environment was generally warmer, resulting in less distinguishable warm skin in the images than when the heaters were

off. To overcome this, the gray values in all the thermal images were mapped to the individual minimum and maximum temperatures to maximize contrast within the dataset.

Image Registration

Due to different fields of view and slightly different capture angles, the RGB and LWIR images required registration to align the modalities. In addition, variations in camera distance and angle required separate registrations for each infant image. Six corresponding points were manually labeled across five image pairs for each recording, resulting in a total of 30 points. These points were distributed evenly across the image to avoid bias towards specific areas. The open incubator and camera movement between measurements required new correspondences to be found for each registration.

Homogeneous coordinates were used to register the RGB and ground truth images to the LWIR image. This approach allows for a single 3×3 transformation matrix. Figure 3.2 illustrates the image transformation.

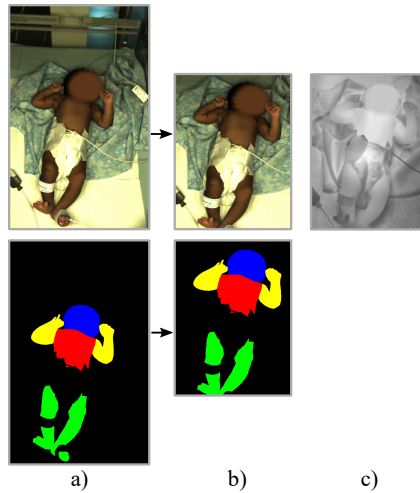


Abbildung 3.2: a) Manually labeled ground truth for the RGB image; b) Transformed RGB image and Ground Truth; c) Corresponding LWIR image.

While libraries such as OpenCV are commonly used for such tasks, training the transformation matrix with L1 loss gave superior results. L1 loss provides

better tolerance for outliers within the image data. However, the limited number of manually labeled points requires protection against overfitting. Therefore, the L1 norm was also used for regularisation. This allows for slight variations in the transformation matrix while maintaining a regularisation effect.

The Root Mean Square Error (RMSE) between all transformed RGB points and their corresponding LWIR points was calculated for both spatial dimensions in a separate test set in order to assess the quality of the image registration. This resulted in an RMSE of $(X, Y) = (6.3 \text{ px}, 10.5 \text{ px})$, which is equivalent to $(X, Y) = (3.15 \text{ mm}, 5.25 \text{ mm})$. For comparison, the OpenCV transformation matrices function was also tested and gave a significantly higher RMSE of $(X, Y) = (15.6 \text{ px}, 26.3 \text{ px})$ or $(X, Y) = (7.8 \text{ mm}, 13.15 \text{ mm})$. However, there is still room for improvement in terms of registration accuracy. Although the camera perspectives varied only slightly during the shooting, some skin areas could be obscured by clothing from one perspective but not from the other. In addition, the manually selected points used to compute the transformation matrices may not be perfectly aligned for both modalities. As a result, the label masks, created from the RGB images, may not perfectly align with the transformed LWIR images. This is a disadvantage when training a network on LWIR images compared to RGB images, or when fusing information from both modalities.

3.2.2 Network Architecture

Body part segmentation of infants is a semantic segmentation task in the medical field, with limited publicly available data. This results in manually labeled small datasets. A U-net architecture was therefore chosen as it was specifically designed to address this type of problem. The success of U-Net in various medical image segmentation and fusion tasks has been well documented [ZZP⁺18]. This neural network is based on an encoder-decoder structure. Here, the encoder extracts features from the input image, while the decoder combines these features to generate the segmentation mask. Furthermore, skip connections within the network connect the encoder and decoder, allowing high-resolution features to be processed in the decoder.

The decoder is made up of five identical blocks. Each block has an up-convolution layer to increase the resolution of the feature map, followed by a 3×3 convolution layer to further improve the features. Batch normalization and ReLU activation were applied after each convolutional layer. The number of filters in the convolutional layers starts with a maximum of 256 and is halved after each block. In the final layer, a softmax activation function was used that generates a class probability for each pixel in the image. These probabilities represent the likelihood that each pixel belongs to a particular body part class.

Finally, these probabilities are converted into a final segmentation mask using the argmax function. This mask assigns each pixel to the class with the highest probability.

Backbone

The performances of different backbones were compared during testing to identify the optimal encoder for the U-Net architecture (see Section A.1 of Appendix). As the DenseNet-121 [HLVDMW17] gave the best result with the least number of parameters (7 million) and similar results were also obtained in the literature for other segmentation tasks [LHKS19], it was chosen as the encoder for this chapter. Unlike the widely used ResNet architecture, DenseNet uses concatenation rather than addition. The network is built using Dense Blocks and Transitional Layers. Within a Dense Block, feature maps maintain their dimensions while the total number of features increases according to a predefined growth rate. Figure 3.3 illustrates a dense block structure with 4 layers and a growth rate of 32.

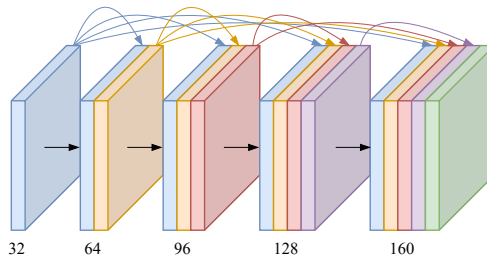


Abbildung 3.3: Dense Block with four layers and a growth rate of 32, as proposed by [HLVDMW17].

Each layer of a Dense Block has six successive operations:

- Batch normalization
- Rectifier Unit (ReLU)
- 1x1 Convolution (Reduction to 128 filters)
- Batch normalization
- Rectifier Unit (ReLU)

- 3x3 Convolution with filter size determined by the growth rate

The layers between the Dense Blocks are called Transitional Layers and are used to downsample the dimension and then apply Batch normalisation, ReLu activation, 1x1 convolution and finally 2x2 average pooling. This backbone architecture has several advantages. First, connecting each layer directly to all subsequent layers via skip connections reduces the problem of vanishing gradients. This ensures that gradients can flow more easily during training, allowing the network to learn effectively from deeper layers. Furthermore, the skip connections also improve feature propagation, as each layer receives not only its own feature maps but also those from previous layers. In this work, a growth rate of 32 was used for the encoder, resulting in an efficient model with only seven million trainable parameters. The decoder structure adds another seven million parameters, bringing the total number of parameters for the proposed network to 14 million.

Fusion Architecture

Unlike the unimodal networks that process either LWIR or RGB images through a single encoder branch, hybrid fusion networks adopt another architecture. First, a simple fusion model is proposed, that uses two separate encoder branches, one for each modality (LWIR and RGB). Features extracted by these branches are then added and fed into the decoder for joint processing. This pixel-wise addition serves a dual purpose: it combines information from both modalities while also minimizing the number of trainable parameters. Figure 3.4 illustrates the proposed simple fusion network architecture.

In addition to simple fusion, a confidence fusion approach was developed and evaluated, which was adopted from Frigo et al. [FMGW22]. This approach takes into account the fact that the quality of the features provided by the individual modality encoders can vary depending on several factors, such as the illumination, the temperature of the infant, and the temperature in the incubator. For example, RGB features may be more informative for segmentation in high-illumination scenarios due to richer texture details compared to single-channel LWIR images. Conversely, LWIR images might provide better features for segmenting classes with low visibility in RGB images under low-illumination conditions. Simply combining features with equal weights could hinder segmentation performance if one modality produces unreliable features. To address this issue, a confidence-based weighting scheme was incorporated.

The confidence weight matrix C is derived from the final decoder layer of the single modality networks after softmax activation. This layer outputs probabilities for each class at each pixel location. An entry C_{ij} of the confidence

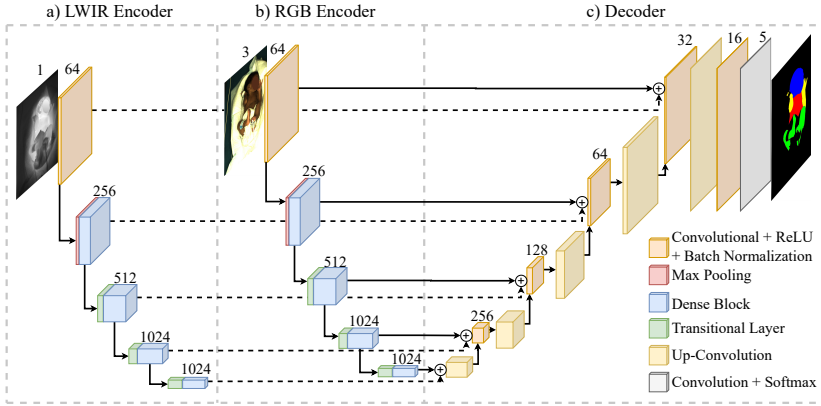


Abbildung 3.4: The simple fusion U-Net architecture with a) an LWIR Encoder based on a DenseNet 121; b) an RGB Encoder based on a DenseNet 121; c) the decoder fusing the outputs of both encoders;

matrix represents the highest probability value at pixel (i, j) . Essentially, it reflects the network’s confidence about its prediction for that specific pixel. The confidence weight matrix is then downsampled to match the spatial dimensions of the encoder features. Weighting is achieved by performing element-wise multiplication between the corresponding entries in the feature maps and the weight matrix. Following this weighting step, the features are fused by simple addition and fed into a common fusion decoder for final processing. A U-Net with confidence fusion is visualized in Figure 3.5.

Implementation

All networks were implemented in Python using the Keras library. The Adam optimizer, proposed by Kingma et al. [KB14], was chosen for optimization using the default Keras parameters. Adam builds on gradient descent by incorporating momentum. This concept helps the network to avoid getting stuck in local minima, which are shallow improvements that do not represent the optimal solution. Momentum achieves this by taking into account not only the current gradient of the loss function but also the gradients from previous training steps when updating the network weights. This allows the network to find the global minimum more effectively [Cho18]. In addition, the standard cross-entropy function was used as the loss function. For the confidence fusion network, the loss is calculated for both the final decoder output and the output of the single

modality models and then equally weighted.

3.2.3 Data Augmentation and Transfer Learning

Due to the limited size of the Chennai dataset, different augmentations were used to artificially expand it and thus improve the generalization ability of the neural networks. A combination of five augmentations was selected after extensive testing. For each augmentation, the maximum degree to which the augmentations are applied (e.g. the maximum scaling factor or the maximum rotation angle) and the probability (p) of application are specified separately:

- The images are randomly **rotated** by a maximum rotation angle of 30° ($p = 0.9$). In the task at hand, rotations can be helpful as they cover a wider range of angular positions of the limbs.
- Randomly **scaling** images up or down by a factor between 0.8 and 1.2 ($p = 0.9$) may improve generalization, as the camera may not always be at the same distance from the infant.
- The images are randomly **shifted** in the horizontal and vertical directions ($p = 0.9$, relative shift limit = 0.2). Shifting an image may improve generalization, although individual filters in a convolutional neural network are invariant to translation. As the infants can move, individual body parts may leave the area captured by the camera. This in turn can be simulated by shift augmentation.
- Horizontal **flipping** of the image ($p = 0.5$), since left/right body parts were not distinguished.
- **Coarse Dropout** of rectangular regions in the image (max. holes = 12, min. holes = 4, min. height = 80, min. width = 80, max. height = 100, max. width = 100, $p = 0.8$). This method proposed by DeVries et al. randomly masks out rectangular regions, which may help to improve the robustness of the neural network [DT17].

In addition to augmentation, transfer learning was applied to overcome the challenges of the small Chennai dataset. Initially, encoder weights trained on the ImageNet data set were used. This dataset, containing over 1.4 million images across 1,000 object classes [RDS⁺15], provided a strong basis for feature extraction, although it was designed for image classification rather than segmentation [HWW⁺19], [ZZP⁺18].

Subsequent pre-training focused on body part segmentation tasks using datasets specifically designed for adults. Hoog Antink et al. combined the Freiburg

Sitting People dataset [OV^B+16] and the Pascal-Person-Part dataset (a subset of VOC 2010) [CML⁺14, EvW⁺10] into the Pascal-Freiburg dataset, containing in total 3733 labeled images. Additionally, the much larger Crowd Instance-level Human Parsing (CIHP) dataset with 38280 labeled images was incorporated for pre-training [GLL⁺18]. An example of these adult datasets visualized with their ground truth labels is illustrated in Figure 3.6.

For both the Pascal-Freiburg and CIHP datasets, the class labels were merged to match the Chennai dataset: head, torso, arm, leg, and background. Notably, the Pascal-Freiburg dataset doesn't distinguish between bare skin and clothed body parts, while the infants in the Chennai dataset typically have bare arms and legs. This difference may be minimal for the legs, but significant for the torso, which is often covered by clothing. However, transfer learning allows for some class mismatches between the pre-training data and the target task.

Since the adult datasets contain only RGB images, they were transformed into a grayscale format for pre-training the LWIR model, as no annotated dataset of adult LWIR images is publicly available.

Fusion Pre-Training

Several pre-training strategies were investigated to optimize the training of the fusion networks. The simplest approach involved pre-training all weights on the CIHP dataset. However, this may not be ideal since the LWIR encoder was pre-trained on grayscale images. In total, four different approaches were tested.

1. Simple Fusion: Pre-train all weights on the CIHP dataset.
2. Simple Fusion: Use the already optimized encoder weights of the fully trained unimodal networks and decoder weights from pre-training on the CIHP dataset.
3. Confidence Fusion: Pre-train all weights on the CIHP dataset.
4. Confidence Fusion: Use the already optimized weights of the fully trained unimodal networks and fusion decoder weights from pre-training on the CIHP dataset.

3.2.4 Evaluation Method

Five-fold cross-validation was used to ensure a robust evaluation of the neural network's performance. The entire dataset of 600 RGB and 600 LWIR images was divided into five folds, each containing 90 RGB and 90 LWIR images. An

independent test set containing the remaining 150 RGB and LWIR images was also created for final evaluation (see Table 3.1 for an overview).

Tabelle 3.1: Split of the Chennai dataset.

Fold	1	2	3	4	5	Test
Number of Recordings	3	3	3	3	3	5
Total image pairs	90	90	90	90	90	150

Importantly, all images from a single recording were assigned to the same fold to maintain consistency within recordings. This resulted in five complete recordings being reserved for final evaluation using the test set. One exception to this split strategy is folds 2 and 4. These folds each contain two separate recordings of the same infant. However, since these recordings were taken on different days with variations in clothing and background, they were assigned to separate folds.

Hyperparameter tuning, including augmentations, network architecture, and pre-training strategies, was performed only during the cross-validation process. The final models were evaluated only once on the held-out test set to avoid the "peeking effect" [Kun14]. This effect occurs when iterative model revisions lead to unintentional tuning of hyperparameters on the test data itself.

To compare the performance of different neural networks, the well-established Intersection over Union metric was used. This metric assesses the overlap between the area assigned to class c by the ground truth mask ($y_{\text{gtruth},c}$) and the area predicted by the network ($y_{\text{prediction},c}$). The IoU is calculated by dividing the area of their intersection by the area of their union:

$$IoU_c = \frac{|y_{\text{gtruth},c} \cap y_{\text{prediction},c}|}{|y_{\text{gtruth},c} \cup y_{\text{prediction},c}|} \quad (3.1)$$

Notably, the IoU can range from 0 (does not overlap) to 1 (both are exactly the same). Furthermore, the mean IoU (mIoU) indicates the accuracy averaged over the four classes head, torso, arms and legs, excluding the background class.

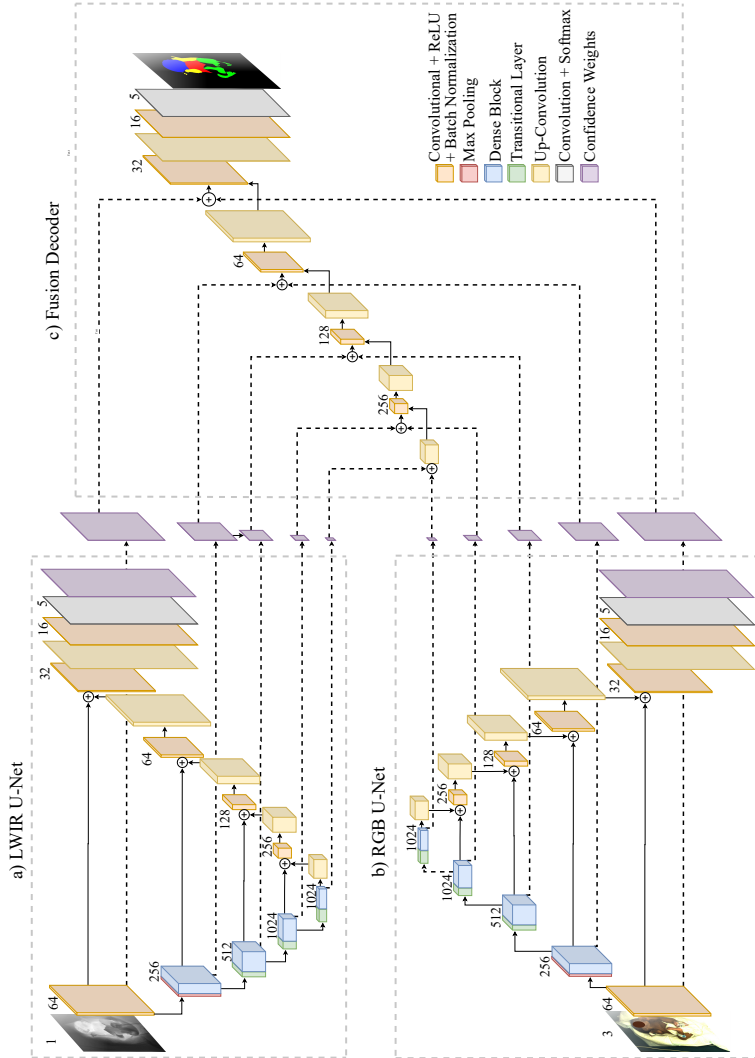


Abbildung 3.5: The confidence fusion U-Net architecture with a) an LWIR U-Net; b) an RGB U-Net; c) the decoder fusing the weighted features of both encoders.

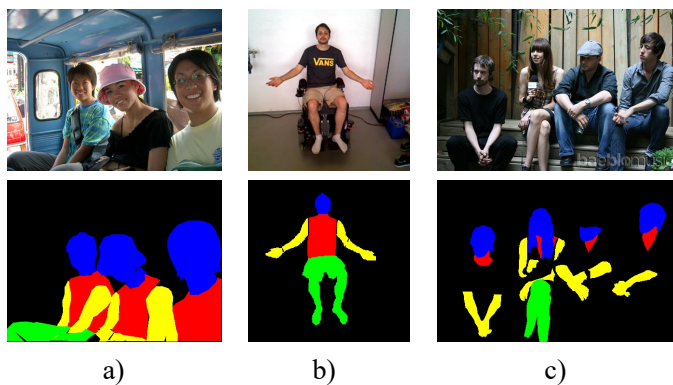


Abbildung 3.6: Examples from a) Pascal-Person-Part Dataset [CML⁺14]; b) Freiburg Sitting People dataset [OV^B+16]; c) Crowd Instance-level Human Parsing dataset [GL^L+18].

3.3 Results

Both unimodal and fusion networks were first optimized using the presented transfer learning and augmentation methods. These optimized models were then evaluated on the held-out test data, incorporating all previous results.

For each fold during cross-validation, training ran for 30 epochs with a batch size of 2, a learning rate of 0.0001, and a scheduled learning rate decay.

3.3.1 Hyperparameter Optimization

RGB Model

First, the influence of pre-training transfer learning on the Pascal-Freiburg and CIHP datasets was investigated to optimize the RGB model. Simultaneously, the effect of data augmentation was investigated. Figure 3.7 portrays the results of the different classes as box plots.

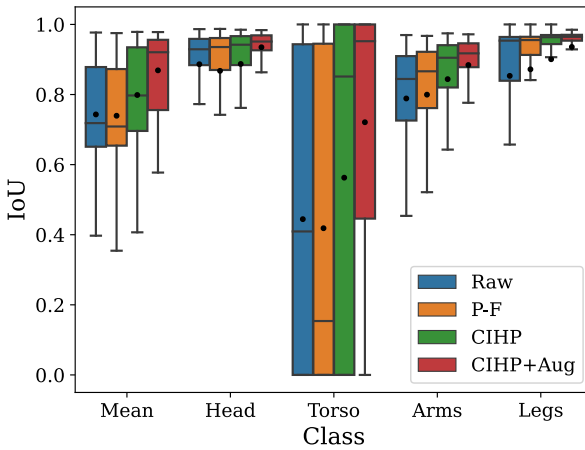


Abbildung 3.7: Box plot for cross-validation of the RGB model with 1) only the raw Chennai dataset (blue); 2) Pascal-Freiburg (P-F) pre-training (orange); 3) CIHP pre-training (green); 4) Data Augmentation and CIHP pre-training (red).

The experiments demonstrate the effectiveness of pre-training on a larger dataset in improving segmentation performance on the Chennai dataset. As depicted in Figure 3.7, models pre-trained on CIHP achieved higher IoU scores for all body parts compared to the model trained only on the Chennai dataset.

On the contrary, pre-training on the Pascal-Freiburg dataset did not result in any improvement, and, in fact, resulted in lower performance for head and torso segmentation.

Overall, the CIHP pre-trained model performed significantly better on all body parts than the Pascal-Freiburg pre-trained model. While the improvement for head, arms, and legs was small, the CIHP dataset significantly improved torso segmentation. In addition, the standard deviation of the IoU scores decreased compared to the model without pre-training, indicating improved stability. Due to these superior segmentation results, CIHP was selected as the sole pre-training dataset for subsequent experiments.

The combination of transfer learning and data augmentation resulted in an even higher mean IoU. Data augmentation improved segmentation performance for all body parts and reduced the standard deviation across folds, resulting in overall good segmentation of infants using only RGB images. However, torso segmentation remained suboptimal with a mean IoU of 0.65.

LWIR Model

Since optimization of the RGB model has already demonstrated superior performance by pre-training on the CIHP dataset, transfer learning for the LWIR model was examined using only a grayscale version of the same dataset. In total, four models were evaluated. Two of these models were trained only on the Chennai dataset, with and without data augmentation. Additionally, two models were pre-trained on the CIHP dataset, also with and without augmentation. The results, as illustrated in Figure 3.8, compare the class performance of the four different models.

Despite the fact that the pre-training was performed only on a grayscale version of the CIHP dataset, the transfer learning resulted in an improvement in IoU of almost three percentage points compared to training only on LWIR images from the Chennai dataset. Including data augmentation further improved the training process, resulting in a mIoU of 0.67.

Comparing the models with augmentation, transfer learning provided a significant improvement of 5 percentage points (from 0.62 to 0.67) over no pre-training. Similar to the RGB model, torso segmentation gave inferior results (IoU of 0.56), while the remaining classes demonstrated good segmentation results, achieving IoUs between 0.66 and 0.76.

Fusion Models

In the following, the modality fusion model is evaluated. Since data augmentation resulted in great improvements for the unimodal networks, it was used

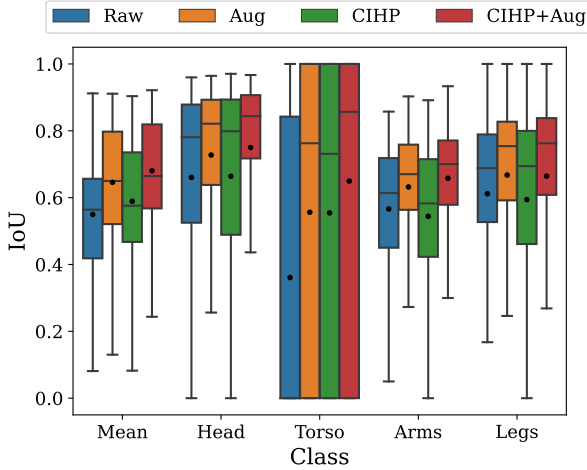


Abbildung 3.8: Box plot for cross-validation of the LWIR model with 1) only the raw Chennai dataset (blue); 2) Data augmentation (orange); 3) CIHP pre-training (green); 4) Data Augmentation and CIHP pre-training (red).

for all training of the fusion model. To optimize the combination of RGB and LWIR images, the four different pre-training strategies are evaluated. The results in Figure 3.9 illustrate the comparison of the classes with box plots.

As can be seen in Figure 3.9, both simple fusion approaches, using either CIHP pre-trained weights or a combination of unimodal encoder weights and a CIHP pre-trained decoder, achieved excellent results with a mean IoU of 0.87. However, the simple fusion model with unimodal encoder weights demonstrated a more consistent performance across body parts, particularly for the torso class. In comparison, both confidence fusion experiments resulted in a slightly lower mean IoU of 0.86. Furthermore, the initial weight selection did not have a significant impact on the results.

3.3.2 Evaluation on Test Set

Based on the results of the hyperparameter optimization, the three final models were then trained on all training data and evaluated on the test set. The pre-training was initially performed on the CIHP dataset for RGB models and on its grayscale version for the LWIR model. For the final evaluation of the simple fusion model, the encoder weights from both unimodal models were used. The decoder weights were taken from the pre-training performed on the

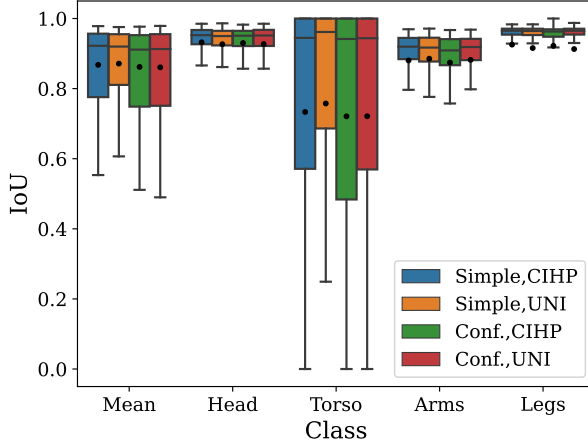


Abbildung 3.9: Box plot for cross-validation of fusion models with Data Augmentation and different fusion pre-training strategies (UNI = weights taken from unimodal networks).

CIHP dataset. Since the confidence fusion model did not show any improvement over the simple fusion model during hyperparameter optimization, it was not investigated further. Figure 3.10 provides the final evaluation results for the three networks as box plots.

Again the fusion model performed best with an IoU of 0.87, closely followed by the RGB model with 0.84 mIoU. However, the LWIR model produced worse segmentation results with an mIoU of 0.75. Looking at the individual classes, the head was segmented most accurately by all three networks, followed by the arms and legs. Similar to the cross-validation results, the torso had the lowest accuracy. In particular, there were significant outliers, resulting in a substantial difference between the median and mean IoUs. For example, the fusion model had a median IoU of 0.90 but a mean IoU of 0.69. Overall, the fusion model consistently achieved comparable or better results than the RGB model across all classes, while the LWIR model lagged behind by 3 to 15 percentage points. For illustration, Figure 3.11 provides qualitative examples of segmentation results.

An example of a superior result achieved by the fusion model is depicted in the upper row of the figure. While the left leg is not detected by the RGB model and the LWIR model has segmentation problems with the torso, the fusion model successfully segments both legs and the torso. Moving to the middle row

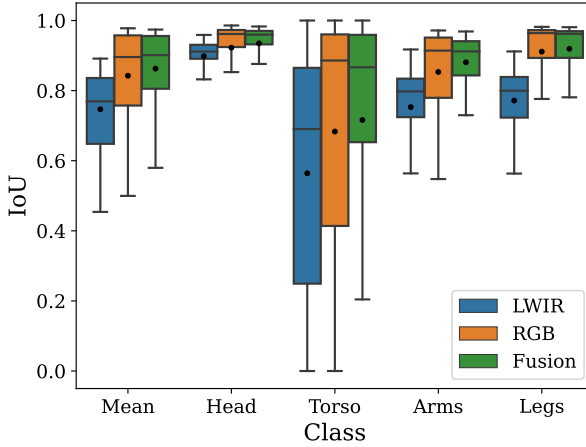


Abbildung 3.10: Box plot for the evaluation of the three different networks on the test set (With pre-training and augmentations), mean is marked with a dot.

of Figure 1, an example of the superior performance of the RGB model is given. Of the three models, it is the only one that accurately segments most of the head and arms. Finally, in the bottom row, the LWIR model outperforms the rest. Unlike the RGB and fusion models, the LWIR model effectively segments the torso skin, yielding a satisfactory result.

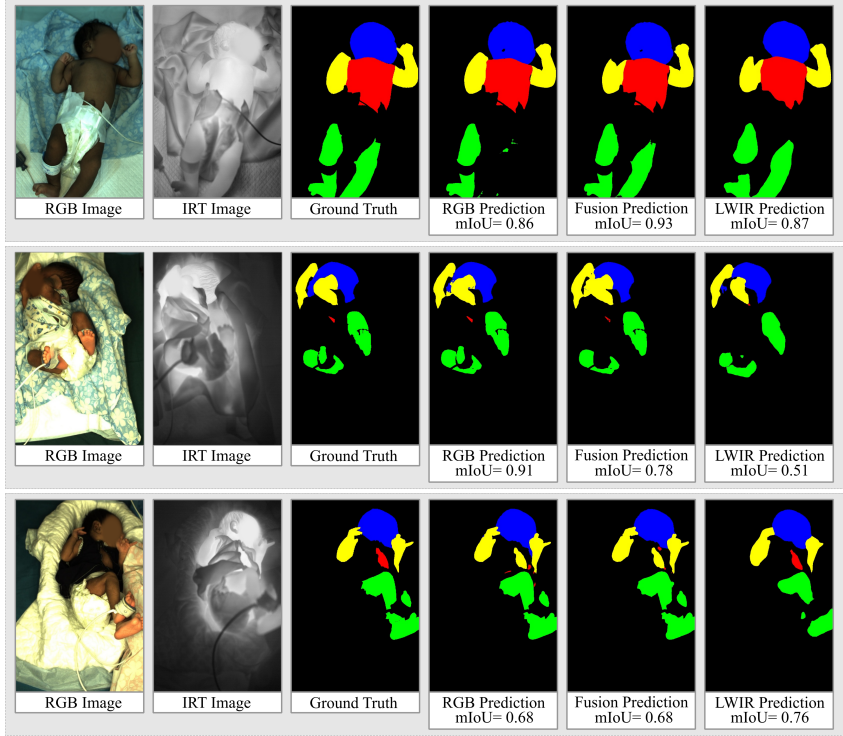


Abbildung 3.11: Qualitative results of the three models evaluated on the test set.

3.4 Discussion

3.4.1 Hyperparameter Tuning

The cross-validation results indicate that transfer learning produces superior results compared to networks trained only on the Chennai dataset. Furthermore, the CIHP dataset proves to be more suitable for infant body part segmentation than the previously used Pascal-Freiburg dataset as proposed by Hoog Antink et al. [AFP⁺20]. Despite the limitation that legs are only visible in about every fourth image of the CIHP dataset, this drawback is compensated by the tenfold larger size of the dataset compared to the Pascal-Freiburg dataset. Notably, the corresponding segmentation results for legs outperformed

those for arms and still surpassed the results of the model pre-trained on the Pascal-Freiburg dataset.

Unlike Asano et al. [AHH⁺22], this work also used transfer learning to train the neural network on LWIR images. Although in this case transfer learning was performed on a grayscale version of the CIHP dataset, it significantly improved the results. This finding suggests that cross-modality transfer learning is feasible and beneficial for semantic segmentation and body part segmentation. However, a larger infant/adult LWIR image dataset would be desirable.

In addition, the employed data augmentation strategy improved the segmentation results for all trained neural networks, resulting in improvements between 5 and 10 percentage points, depending on the model. Interestingly, Hoog Antink et al. [AFP⁺20] observed a slight degradation in results when applying data augmentation. They speculated that their network might generalize better with augmentation, but achieving good results on the validation data from their relatively small dataset (563 RGB images in total) did not necessarily require augmentation.

Furthermore, the Chennai dataset is only slightly larger than the dataset used by Hoog Antink et al. and contains only small variations in color, brightness, and camera position. This limited diversity may introduce biases, particularly with respect to skin color. To determine the influence of data enhancements such as color/brightness/saturation variations on the skin, additional experiments were conducted. The result was the same accuracy as without any skin color augmentation, an indication of skin color bias in the Chennai dataset. Overall, when using the neural network in real-world scenarios or with new data with greater variations in skin color and brightness, data augmentation (including color and brightness variations) will be essential.

In addition, transfer learning was performed on grayscale RGB images from the CIHP dataset to improve the LWIR model. Given the lack of large thermographic databases with labeled body parts, this approach was the most feasible option despite the differences in modalities. In the future, the generation of additional thermography data or even the creation of artificial image data could further improve transfer learning.

Regarding the choice of backbone architecture, a DenseNet121 was chosen for the neural network. Although DenseNets are sometimes less memory and speed-efficient than other network structures, this trade-off is acceptable in the context of non-contact temperature measurement. However, exploration of alternative backbones, such as the Repvgg proposed by Ding et al. [DZM⁺21], should be considered for segmentation tasks. Such investigations could lead to improved inference times and enable applications such as camera-based vital signs measurement.

3.4.2 Image Fusion

The results of this chapter are consistent with the findings of Farahnakian et al. [FH20], demonstrating the superiority of fusion models over unimodal approaches. In addition, similar to their study, RGB models showed better performance compared to LWIR models.

However, in contrast to Farahnakian et al., the presented fusion model achieved only slightly better results than the unimodal ones, with some classes even showing slightly worse segmentation. This result is unexpected since fusion models typically benefit from the combined information of both modalities. A possible explanation lies in the different scenarios studied. Farahnakian et al. [FH20] focused on maritime ship detection using RGB and LWIR images with significant distances between targets and cameras (tens or hundreds of meters). In such cases, perspective discrepancies due to separate cameras are less significant compared to situations with closer proximity, as seen in this study where the cameras are a maximum of two meters away from the children. Similar observations can be found in other multimodal studies [HWK⁺17, LW19, SZL19, WZA⁺05, ZCVZ15] that focus on landscapes, ship traffic, satellite imagery, and urban scenes, where target distances are considerably larger. In addition, some studies used single cameras capable of capturing both RGB and thermal images simultaneously, further minimizing perspective inconsistencies [HWK⁺17, SZL19].

Another factor influencing the performance difference is the lower contrast between skin and background often observed in the LWIR images of the Chennai dataset. This reduced contrast may pose additional challenges to the segmentation process on this modality. Furthermore, the manual transformation of RGB and ground truth images may not achieve perfect registration. This misalignment may explain why the fusion model in this chapter occasionally performed worse than the unimodal networks. Investigating improved image transformation techniques, such as automatic registration methods using calibration objects, could potentially improve the performance of fusion models in future work.

Interestingly, the confidence fusion approach, despite its sophistication, did not yield an improvement in accuracy over simple fusion. This may be due to the characteristics of the Chennai dataset. The controlled environment during recording with external lighting likely minimized the variation of situations where one modality would provide significantly more reliable features than the other. However, in real-world scenarios with varying and uncontrolled lighting conditions, confidence fusion may prove more advantageous. For example, in low-light conditions, the information from the LWIR modality may be more reliable due to its independence from visible light. In such cases, the confi-

dence weighting mechanism could automatically prioritize the LWIR features, resulting in improved segmentation accuracy. Exploring the performance of confidence fusion in various real-world settings with uncontrolled lighting and environmental variations is important for future research.

In addition, the concatenation of the extracted features from the different modalities was tested. However, this resulted in the same IoU of 0.87 during the optimization of the hyperparameters. Since the concatenate operation adds six million more weights to the fusion model, the add operation was chosen instead. Other fusion methods (early fusion and late fusion) were also tested, but the results were also inferior.

3.4.3 Class Prediction

Evaluation of all three models showed good overall segmentation results for all body part classes, with the head consistently achieving the highest accuracy. However, a challenge was found in the torso segmentation, where all the networks showed the lowest performance. This difficulty can be attributed to the amount of clothing covering the torso for the majority of infants in the Chennai dataset. In Table 3.2, the average percentage of pixels of each class from the whole image can be seen.

Tabelle 3.2: Classwise percentage of pixels for the Chennai dataset.

Class	Background	Head	Torso	Arms	Legs
Percentage	83.4 %	4.6 %	1.8 %	3.6 %	6.5 %

As shown in Table 3.3, on average the torso skin makes up only 1.8 % of the total image area. In fact, clear visibility of the torso skin was observed in only six of the recorded infants. In the remaining images, the torso was either completely covered by clothing or only small portions of skin were visible. The networks struggled to achieve good segmentation results under these unfortunate conditions.

Figure 3.12 further illustrates this point, showing a clear correlation between the IoU for the torso and the size of the visible skin area. As the visible skin area decreases, the IoU for the torso also decreases. However, it's important to note that for accurate temperature measurement, a certain minimum skin area is required to obtain reliable readings. This means that temperature measurements cannot be reliably extracted from skin patches as small as those observed in some cases in the Chennai dataset.

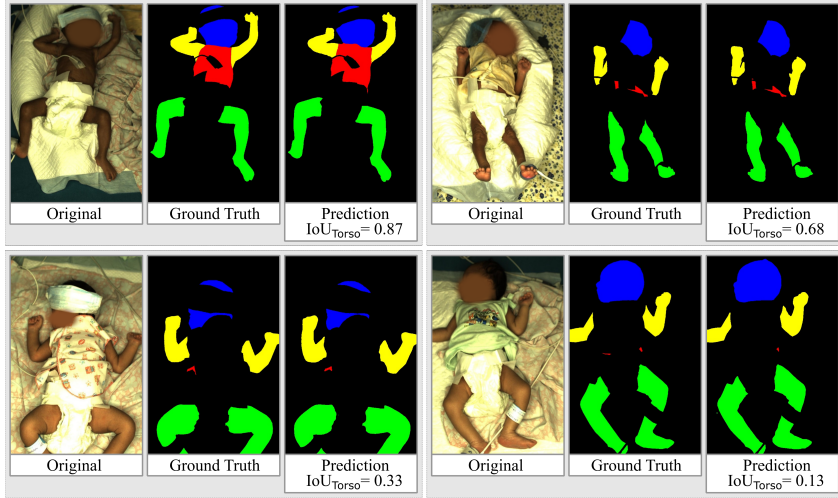


Abbildung 3.12: Examples of images and respective ground truth/prediction for different qualities of torso segmentation.

Therefore, a practical approach to address this challenge is to set a threshold for the minimum size of visible skin areas required for temperature extraction. By implementing this strategy, the impact of suboptimal segmentation results on the torso, particularly in cases with minimal skin visibility, can be reduced.

3.5 Summary

This chapter investigated the development and evaluation of a deep learning model for automatic body part segmentation in infants, with the goal of enabling non-contact temperature monitoring using infrared thermography. Different U-Net based neural network architectures were compared, using either RGB or LWIR images individually, or a fusion approach combining information from both modalities.

A dataset of 600 RGB and 600 corresponding LWIR images of premature infants was manually annotated and used for training and evaluation. Optimization of the individual models showed that transfer learning with the CIHP dataset, in combination with data augmentation, significantly improved segmentation accuracy across all modalities. Interestingly, even cross-modality transfer learning, where the LWIR model was pre-trained on a grayscale version

of the CIHP dataset, showed significant improvements.

Following these results, the final RGB, LWIR, and fusion model were evaluated on a separate test set. The fusion model, which uses a simple addition of features extracted from both modalities, achieved the best overall segmentation performance with a mIoU of 0.87. The RGB model followed closely with an mIoU of 0.84, while the LWIR model lagged slightly behind with an mIoU of 0.75. Across all models, head segmentation consistently showed the highest accuracy, followed by arms and legs. Torso segmentation remained a challenge, especially in cases with minimal skin visibility due to clothing. However, implementing a threshold for the minimum skin area considered for temperature measurements provides a practical solution to overcome this limitation.

Evaluation of different fusion strategies revealed that simple feature addition provided comparable performance to more complex concatenation approaches while requiring fewer parameters. Furthermore, the confidence fusion method, which dynamically weights features based on the model's confidence in each modality, did not lead to significant improvements over simple fusion in this controlled environment. However, it is promising for real-world scenarios with varying lighting conditions and uncontrolled environments, where the confidence weighting mechanism could automatically prioritize the more reliable modality.

In conclusion, this chapter demonstrates the potential of multimodal deep learning approaches for accurate and robust body part segmentation of infants. Future research should focus on expanding the dataset to include different skin tones, body positions, and real-world scenarios, as well as exploring advanced image registration techniques and investigating the performance of confidence fusion in more complex environments.

4 Spatial Temperature Monitoring of Premature Infants

Accurate and reliable monitoring of skin temperature is essential for the thermoregulation of premature infants. Current methods using wired sensors are invasive and prone to error. This chapter presents a novel non-contact skin temperature monitoring system for premature infants using infrared thermography. To mitigate systematic errors caused by factors such as infrared window transmission, reflections from incubator walls and camera drift, a novel temperature correction algorithm was developed and rigorously validated using a blackbody reference. The algorithm achieved a mean absolute error of 0.17°C and a maximum error of 0.83°C compared to a reference thermistor. Body part segmentation, using a U-Net model with transfer learning and data augmentation, gave a mean Intersection over Union of 0.77. Spatial temperature extraction, using a percentile-based method, gave satisfactory accuracy with a mean absolute error of 0.41°C compared to the reference adhesive temperature sensor. Furthermore, the torso and arm regions provided more robust central and peripheral temperature measurements than the head and legs regions. The results demonstrate the reliability of this non-contact system for clinically relevant temperature monitoring in premature infants. Parts of this chapter have already been published in peer-reviewed journals and conferences [VWB⁺22], [VLN⁺25].

4.1 Motivation and Overview

Incubators are vital to create an optimal thermal environment for premature infants. It is essential to monitor and regulate body temperature as it constitutes an important vital sign. In order to ensure the continuous monitoring of the infant's health status, the temperature is typically assessed through the use of thermal skin probes. However, frequent changes of these probes can result in damage to the skin and an increased risk of infection [HHA⁺22a]. Furthermore, multiple sensors are required in order to provide information about the state of an infant's thermoregulation [LP15]. A contactless thermography system has the potential to become a valuable alternative in neonatal care. Long-wave infrared cameras represent a promising tool for this purpose. The capacity to contactlessly measure multiple points simultaneously could offer a more comprehensive temperature monitoring than current standards, as well as a more comfortable experience for the infant, parents and care staff.

Although there has been some research into IRT for premature infants, it has mainly focused on measuring respiratory rate [AHJ⁺11, KVDBVDV⁺14, PYG⁺18, LSM⁺21]. In contrast, IRT for measuring body temperature has not been well studied. In particular, measuring the temperature of premature infants requires high accuracy, which current infrared cameras do not provide.

There have been different approaches to advance this area of research. Abbas et al. investigated influences on IRT in a neonatal clinical setting, including convective incubators. A model-based correction was developed to improve the accuracy of the measurement [AHB⁺12]. However, this study did not use a thermal reference and did not perform direct accuracy measurements with a blackbody in the clinical setting.

In contrast, Hamada et al. [HHA⁺22a] investigated the accuracy of infrared cameras placed inside the incubator for temperature measurements. For this purpose, they used two different uncooled LWIR cameras (Flir A35 and Lepton 3.5), placed inside the incubator, along with a passive and an active blackbody for reference and accuracy measurement. They measured 24 different scenarios with different incubator and blackbody temperatures and two different humidity levels. For compensation, they scaled the difference between the measured object and the measured reference and added an offset and the unscaled flux of the ground truth reference. The scaling and offset factors are determined by solving a regression problem for two of the 24 measurement scenarios. For the calculation of the results, they first performed an outlier detection and discarded these values. This approach reduced the mean absolute error (MAE) for the high-cost A35 to below 0.3 °C in all scenarios. For the low-cost Lepton 3.5 an MAE below 0.3 °C was achieved in only 19 scenarios [HHA⁺22a].

Existing research addressing this specific application has limitations, with studies either focusing solely on accuracy improvements through correction algorithms [AHB⁺12, HHA⁺22a] or focusing only on automated image segmentation without investigating temperature measurements [AFP⁺20, AHH⁺22].

To date, there is no existing literature describing a clinically validated IRT-based approach to automatically measure the temperature distribution of premature infants with high accuracy.

This chapter introduces a novel approach to accurately measure the temperature distribution of premature infants. This approach integrates IRT with a robust algorithm for automated skin segmentation. For this purpose, the segmentation algorithm developed in Chapter 3 was adapted to a new clinical setting. Furthermore, a new temperature correction algorithm was developed and evaluated with an accurate blackbody to ensure reliable measurements with IRT. Finally, the clinical validity of this approach was demonstrated in a validation study.

First, the hardware setup and the recorded dataset of the clinical study are outlined in Section 4.2. Next, in Section 4.3 the novel correction algorithm will be explained, which enables accurate thermography for neonatal incubators. After that, the adapted segmentation algorithm will be presented and evaluated in Section 4.4. Finally, in Section 4.5 the automated temperature extraction pipeline will be validated.

4.2 Clinical Study

The clinical study was conducted in the neonatal intensive care unit of the RWTH Aachen University Hospital (UKA), Aachen, Germany and approved by the ethics committee of the UKA, Aachen, Germany (EK 22-315). Written informed consent was obtained from the parents of all patients.

4.2.1 Hardware Setup

A schematic representation of the hardware setup is illustrated in Figure 4.1.

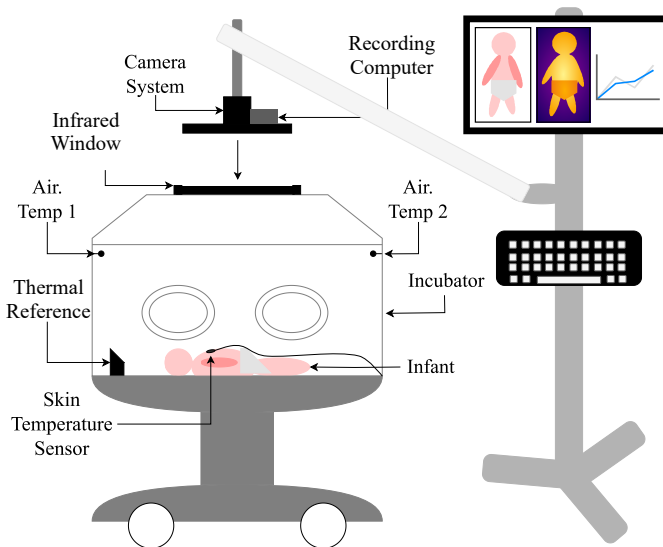


Abbildung 4.1: Schematic Representation of the clinical hardware setup.

All recordings of the clinical study were collected with a Thermocare Vita incubator (Weyer GmbH, Germany). Normally, the standard incubator cover is not transparent to infrared radiation. Therefore, a *ClearIR-4-P* (Iriss) infrared window, which is transparent to long-wave infrared radiation (8-14 μm), has been incorporated into the cover of this research incubator. Inside, two sensors monitored the air temperature. The average value of both was saved as the incubator air temperature. An additional sensor measured the skin temperature of the premature infants, which was attached to their torsos and used as a reference for subsequent temperature extraction. The camera system, with both

an RGB and an infrared camera, was positioned centrally on the top of the incubator. This system, together with the recording computer (Jetson Xavier AGX from Nvidia), was placed inside a custom-made mount, which in turn was fixed to a medical cart.

Figure 4.2a depicts an external view of the real setup with the cameras mounted on the top of the incubator. In addition, Figure 4.2b provides the view from inside the incubator. Here, the infant (mimicked by a neonatal phantom) lies on the mattress and the cameras can be seen at the top of the incubator. The visible lights were switched off during the recordings.

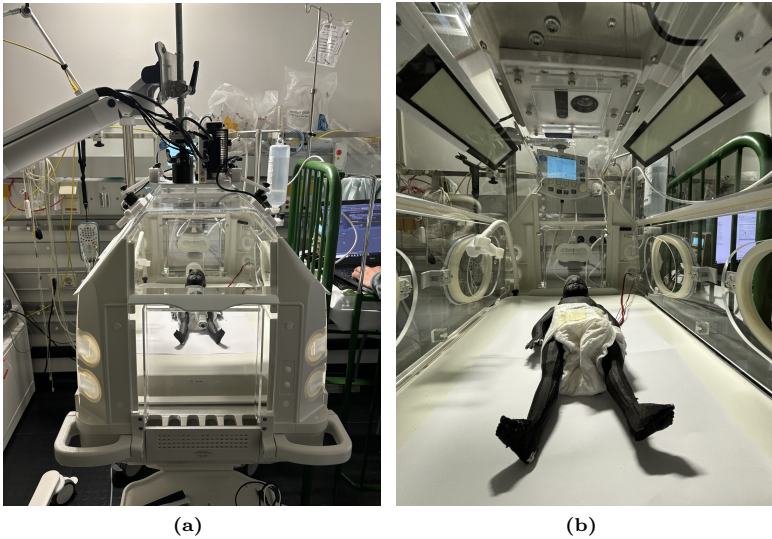


Abbildung 4.2: Hardware setup of the clinical study with a) exterior and b) interior view of the incubator.

The RGB images were captured with a 12-megapixel OAK-D (Luxonis) camera. The LWIR images were taken with a PI 640 camera from Optris. The main properties of this LWIR camera are listed in Table 4.1.

In addition, the thermal reference presented in Section 4.3.1 was placed inside the incubator for later calibration of the infrared camera.

Tabelle 4.1: Properties of the used IRT camera

Property	Optris PI 640
Resolution	640 x 480
Thermal sensitivity	75 mK
Absolute accuracy	$\pm 2^{\circ}\text{C}$
Field-of-view	$90^{\circ} \times 66^{\circ}$
Price	7500 €

4.2.2 Recordings

The dataset for this chapter consists of a total of 14 recordings of 14 different premature infants. Example images of the dataset can be seen in Figure 4.3.



Abbildung 4.3: Example images of the dataset. Top: processed RGB images; Bottom: corresponding, registered LWIR images.

Each recording was captured at a frame rate of one frame per second for both image modalities over a period of five hours. Alongside the images, both the skin temperature measurement of the torso and all incubator temperature data were saved with a recording rate of 1 Hz. An overview of the patients is listed in Table 4.2.

Tabelle 4.2: Overview of all patients in the clinical study.

ID	Gender	Gestat. age (weeks)	Birth weight (kg)	Birth height (cm)	Age (days)	Rec. weight (kg)
PA01	m	31	1.1	38	8	1.05
PA02	m	31	0.8	37	15	1.07
PA03	w	30	1.13	38	11	1.035
PA04	w	29	1.225	40	10	1.25
PA05	w	33	1.38	39	8	1.285
PA06	m	29	1.05	38	17	1.32
PA07	m	29	1.09	35	14	1.23
PA08	m	30	0.965	37	14	1.08
PA09	w	28	0.81	34	26	1.09
PA10	w	31	1.41	41	14	1.41
PA11	w	31	1.39	40	14	1.52
PA12	m	27	1.18	36	20	1.29
PA13	w	28	0.935	35.5	22	1.26
PA14	w	27	0.69	31	38	1.21
Mean	8 × w 6 × m	29.57	1.08	37.11	16.5	1.22

The infants had a gestational age of 27 to 33 weeks at birth and a birth weight of 0.69 kg bis 1.41 kg. At the time the images were taken, they were between 8 and 40 days old and weighed between 1.05 kg bis 1.52 kg. The inclusion criterion for the study was a minimum weight of 1 kg.

4.3 Accurate Thermography for Incubators

Accurate absolute temperature measurement using infrared thermography requires careful consideration of various influences, as discussed in Section 2.2.3. This is particularly crucial when measuring the temperature of a premature infant within an incubator, as the enclosed environment introduces additional complexities to the radiation model. An IRT camera positioned outside the incubator measures the thermal radiation that passes through an IR-transparent window. However, this window, along with other factors within the incubator,

significantly affects the radiation received by the camera. Figure 4.4 presents a schematic representation of the total radiation, received by the IRT camera.

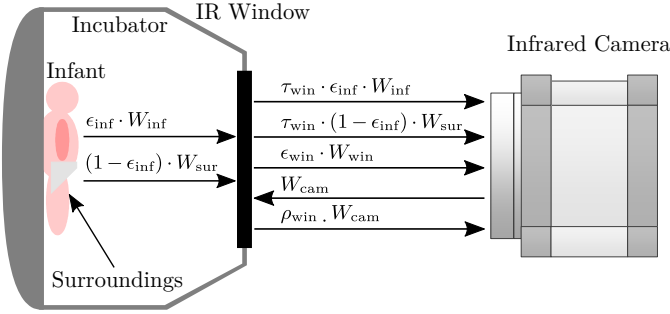


Abbildung 4.4: Total Radiation received by an infrared camera for a neonatal incubator.

The primary challenge is to accurately detect only the radiation from the infant (W_{Inf}). This is because the camera also captures radiation from other sources. The total radiation received (W_{tot}) is a combination of several components:

1. Infant's emitted radiation: The target measurement, W_{inf} , is the thermal radiation emitted by the infant's skin. This radiation is dependent on the skin temperature (T_{inf}) and emissivity (ϵ_{inf}).
2. Incubator Surroundings: A portion of the thermal radiation emitted by the incubator surroundings W_{sur} is reflected by the infant's skin. The amount of reflected radiation depends on the infant's skin reflectivity ($1 - \epsilon_{\text{inf}}$).
3. IR Window: The IR window, although not perfectly transparent to infrared radiation, transmits a large fraction of the infant's and surrounding's radiation, dependent on the window's transmissivity (τ_{win}). Additionally, the window itself emits thermal radiation (W_{win}) based on its temperature and emissivity (ϵ_{win}).
4. Camera Radiation: A portion of the radiation emitted by the camera itself (W_{cam}), which depends on the camera temperature T_{cam} , is reflected by the IR window back towards the camera. This reflected component is dependent on the window's reflectance (ρ_{win}).

Therefore, the total radiation detected by the IRT camera (W_{tot}) can be modeled as the sum of these components:

$$W_{\text{tot}} = \tau_{\text{win}} \cdot \epsilon_{\text{inf}} \cdot W_{\text{inf}} + \tau_{\text{win}} \cdot (1 - \epsilon_{\text{inf}}) \cdot W_{\text{surr}} + \epsilon_{\text{win}} \cdot W_{\text{win}} + \rho_{\text{win}} \cdot W_{\text{cam}} \quad (4.1)$$

This equation forms the basis for understanding the influences on neonatal thermography within an incubator environment when measuring through an infrared window. It highlights the need for a robust calibration approach to accurately determine the infant's skin temperature.

4.3.1 Setup for Calibration

Typically, thermal reference sources are used to calibrate infrared cameras and compensate for influences [SVP⁺20]. In this case, the thermal reference sources are placed in the field of view of the camera. This section presents a setup for temperature calibration using a custom thermal reference inside an incubator, with the objective of improving the absolute accuracy of neonatal temperature measurement with an IRT camera.

A custom-built passive thermal reference was developed for drift/noise compensation of the infrared camera. The setup of the thermal reference can be seen in Figure 4.5.

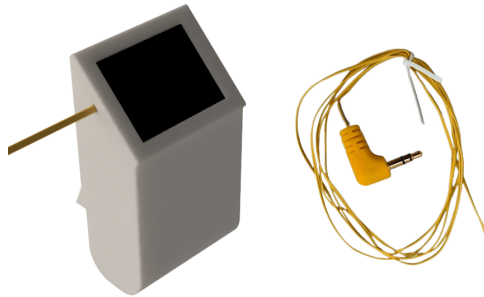


Abbildung 4.5: Custom made passive thermal reference (left) with clinically used temperature sensors (right).

An NTC thermistor (Weyer GmbH) was selected as the accurate reference temperature sensor, which is clinically used to measure the skin temperature of infants. This sensor has several advantages. Besides a high accuracy of $\pm 0.1^\circ\text{C}$,

it can be read out directly from the incubator in the setup, requiring no additional hardware. The sensor was then integrated into a silicone housing, which was created by pouring *SILIXON10* (silikonfabrik.de, Ahrensburg, Germany) into a 3D-printed casting mold. In addition, the top of the silicone housing was covered with a black coated aluminum foil, having a high emissivity of over 98 % (Advanced Coating, Bad Kösen, Germany). The silicone housing was designed to fit precisely into the sides of the incubator, allowing for a stable position during measurements.

The complete setup for accurate thermography for incubators is depicted in Figure 4.6, including a blackbody for accuracy evaluation.

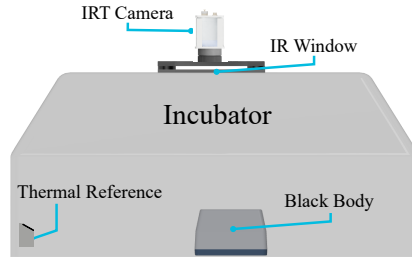


Abbildung 4.6: Measurement Setup including active blackbody and passive thermal reference.

The Thermocare Vita incubator (Weyer GmbH, Kürten, Germany), equipped with the same IR window as in the clinical study, was also used for evaluation measurements. Instead of an infant, the active blackbody "Nightingale Body Temperature Reference" (Santa Barbara Infrared Inc., Santa Barbara, USA) was chosen as a measuring body. The blackbody's emissivity ($>95\%$) is similar to that of human skin, making it a suitable substitute. In addition, a uniform temperature with an accuracy of $\pm 0.15\text{ }^{\circ}\text{C}$ can be set. The connection to the blackbody was established on the hardware side via the micro USB port and on the software side via serial interface. Finally, the IRT camera was fixed on top of the incubator.

4.3.2 Calibration Algorithm

As described before, thermal radiation from the IRT camera reflects off the infrared window, interfering with the measurement. After a warm-up period, this signature produced by the infrared window manifests as a steady-state, additive disturbance superimposed on the image. This disturbance is compensated

for by applying a non-uniformity correction (NUC) using the matrix defined in equation (4.2).

$$T_{\text{NUC, Mat.}}(x, y) = \frac{1}{15} \sum_{n=1}^{15} (T_{\text{hom.,cam}}(x, y)(t_n) - T_{\text{ref,surf.}}(t_n)) \quad (4.2)$$

Here $T_{\text{NUC, Mat.}}$ is the resulting correction matrix, $T_{\text{hom.,cam}}$ is the thermal image of a uniformly radiating calibration surface as measured by the camera, and $T_{\text{ref,surf.}}$ is the true temperature of that surface. Fifteen images were taken and averaged to remove possible noise. The black-coated aluminum foil (selected for the passive thermal reference as described earlier) was attached to a flat 3D-printed surface. This setup served as the radiating calibration surface. This calibration surface was placed in the incubator, which was maintained at 32 °C, for a sufficient duration to reach thermal equilibrium. Thus, T_{ref} is assumed to be equal to the incubator temperature. Since this disturbance is assumed to be steady-state and additive, the IRT measurements can be compensated by:

$$T_{\text{cam, corr.}}(x, y)(t) = T_{\text{cam}}(x, y)(t) - T_{\text{NUC, Mat.}}(x, y) \quad (4.3)$$

Figure 4.7 depicts the calibration surface recorded by the IRT camera (left) and the NUC-corrected IRT image (right).

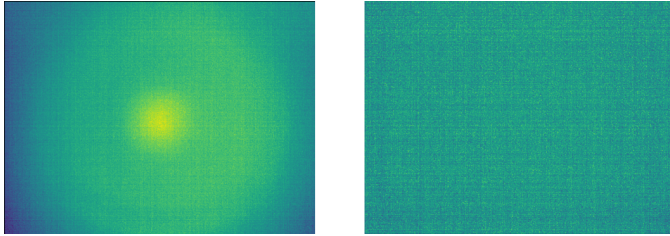


Abbildung 4.7: Calibration surface recorded by IRT camera (left) with NUC correction (right)

The remaining influences on the IRT camera were compensated with a custom correction equation, which is based on the IRT-measured temperature $T_{\text{Temp,IRT}}$, the IRT-measured thermal reference temperature $T_{\text{Ref,IRT}}$ and the sensor-based thermal reference temperature $T_{\text{Ref,Sens}}$.

Therefore, the corrected temperature $T_{\text{Temp,Corr}}$ can be calculated as follows:

$$T_{\text{Temp,Corr}} = a_0 + a_1 \cdot (T_{\text{Temp,IRT}} - a_2 \cdot (T_{\text{Ref,IRT}} - T_{\text{Ref,Sens}})) + a_3 \cdot T_{\text{Ref,Sens}} \quad (4.4)$$

with $a_0 - a_3$ being fitted coefficients. The correction algorithm has been designed to address the combined effects of drift, environmental influences, and the incubator air temperature. The term $T_{\text{Temp,IRT}} - a_2 \cdot (T_{\text{Ref,IRT}} - T_{\text{Ref,Sens}})$ aims to reduce the impact of both temporal drift in the IRT camera readings and the influence of the surrounding incubator environment, specifically radiative effects from the incubator walls and variations in the window's thermal properties. The difference $T_{\text{Ref,IRT}} - T_{\text{Ref,Sens}}$ quantifies the error introduced by the IRT camera's measurement of the reference temperature. Scaling this error by a_2 and subtracting it from the raw IRT temperature ($T_{\text{Temp,IRT}}$) effectively compensates for these combined effects.

Figure 4.8 provides the IRT-measured temperature $T_{\text{BB,IRT}}$ and the reference-corrected temperature $T_{\text{BB,DriftCorr}}$ of a blackbody (heated to 37 °C), the IRT-measured thermal reference temperature $T_{\text{Ref,IRT}}$ and the sensor-based thermal reference temperature $T_{\text{Ref,Sens}}$.

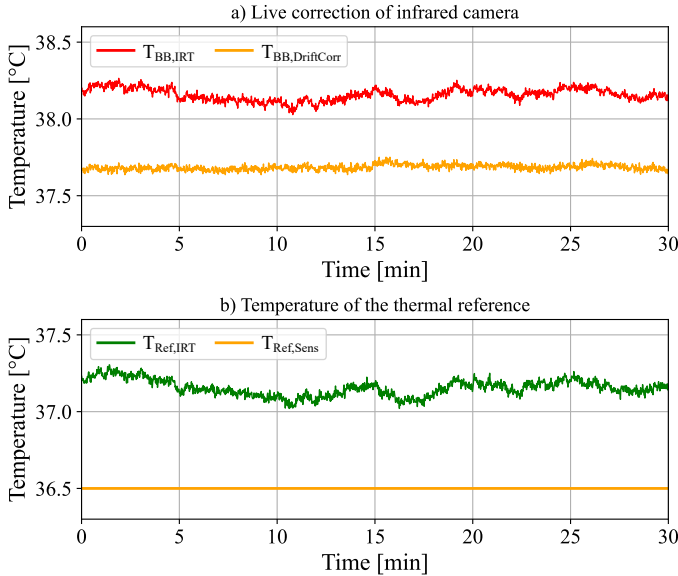


Abbildung 4.8: Drift correction using the thermal reference with a) the IRT-measured and corrected blackbody temperature; b) IRT-measured and sensor-based thermal reference temperature;

It is evident that the IRT-measured temperature of the blackbody exhibits the same fluctuations as the IRT-measured reference temperature. Using the

live reference correction $T_{\text{BB,DriftCorr}} = T_{\text{BB,IRT}} - a_2 \cdot (T_{\text{Ref,IRT}} - T_{\text{Ref,Sens}})$, the corrected temperature is obtained (yellow line), with the fluctuations being significantly reduced. Moreover, the slight attenuation of the infant's thermal radiation and the radiation of the IR window itself can be compensated for by a gain-offset correction with a_0, a_1 .

Finally, the incubator's air temperature significantly influences the IRT measurements. The term $a_3 \cdot T_{\text{Ref,Sens}}$ incorporates this influence. The precise sensor-based temperature of the thermal reference $T_{\text{Ref,Sens}}$ serves as a robust approximation for the incubator air temperature. The scaling factor a_3 is used to precisely quantify the contribution of the incubator air temperature to the final corrected temperature.

4.3.3 Evaluation Measurements

A total of 26 recordings, each lasting seven hours, were conducted, to provide a robust evaluation of the infrared camera and correction algorithm performance. To prevent overfitting to a specific location, the blackbody's position within the incubator was varied for each recording. The positions were determined using a pre-defined grid (Figure 4.9), resulting in a total of 24 different positions. Two additional measurements were taken in a central position, as premature infants often lie in the central area of the incubator.

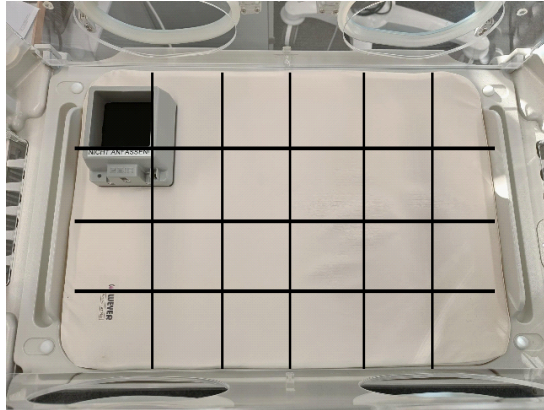


Abbildung 4.9: Measurement grid including active blackbody.

For each recording, the incubator temperature T_{Inc} was varied between 31 °C and 37 °C, while the blackbody temperature $T_{\text{BB,real}}$ was varied between 35 °C

and 40 °C. This resulted in each recording lasting approximately seven hours, as depicted in Figure 4.10.

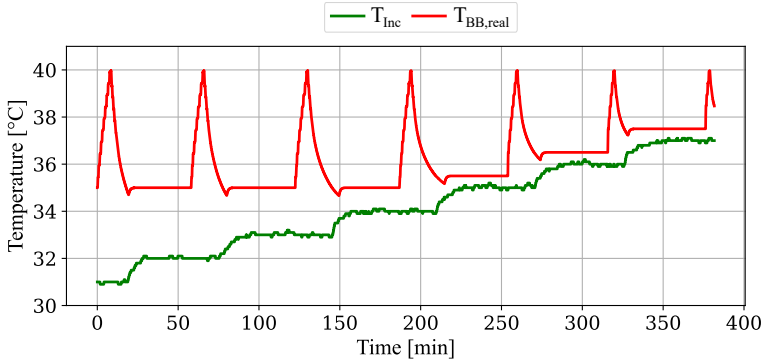


Abbildung 4.10: Temperature settings for laboratory measurements.

This range of temperatures captured all clinically relevant combinations of incubator and infant skin temperatures, including cases of both hypothermia and hyperthermia. All 26 recordings were split into five folds for a five-fold cross-validation. At the same time, it was ensured that all data of one recording was only assigned to one fold.

4.3.4 Validation Results and Discussion

The calibration coefficients were determined using the Levenberg-Marquardt algorithm with a least-squares cost function, applied to the dataset described in Section 4.3.3. Performance evaluation of the resulting calibration was conducted using five-fold cross-validation. First, a qualitative validation involved comparing the corrected IRT blackbody temperatures to the reference sensor readings. Figure 4.11 presents an example of this comparison, showing the IRT-measured $T_{BB,IRT}$ and corrected $T_{BB,Corr}$ blackbody temperatures, along with the incubator air temperature T_{Inc} and the true blackbody temperature $T_{BB,real}$ for a single measurement.

The raw IRT temperature (red line) deviates significantly from the true blackbody temperature (dashed yellow line). It is also evident that alterations in incubator air temperature (illustrated by the green line) have an impact on the IRT measurement. However, the proposed correction algorithm (blue line) effectively compensates for these influences, closely tracking the true blackbody temperature (dashed yellow line).

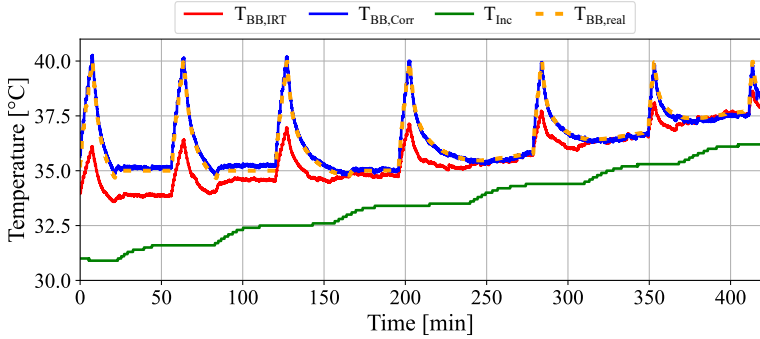


Abbildung 4.11: Qualitative results of temperature correction for one measurement.

In addition, quantitative validation was performed using the results of the five-fold cross-validation. Table 4.3 presents the mean absolute error (MAE), maximum error (MAX), and standard deviation (STD) of the IRT temperature measurements, both before (raw) and after (corrected) applying the calibration algorithm.

Tabelle 4.3: Qualitative results of temperature correction for one measurement.

Metric	Raw	Corrected
Mean absolute error	0.91	0.17
Max. Error	4.85	0.83
Standard deviation	0.79	0.12

The calibration method described in this chapter substantially improves temperature measurement accuracy. The measurement error is reduced by 81.3% compared to the raw measurement. The maximum error is also reduced from 4.85°C to 0.83°C, a reduction of 83%.

The comprehensive data set includes a wide range of incubator and black-body (infant) temperatures, as well as different blackbody positions within the incubator. Furthermore, the implementation of cross-validation reduces the risk of overfitting the correction algorithm to the calibration data.

4.4 Body Part Segmentation

This section first presents the dataset used for body part segmentation, including the steps for pre-processing and registering the modalities. Then the evaluation method will be explained, including the employed deep learning architecture. Finally, the results of the segmentation will be outlined and discussed.

4.4.1 Dataset

A total of 959 image pairs were manually selected from all recordings (see Section 4.2). Special attention was paid to a high diversity of images in terms of posture and position, size of the infant and brightness. Between 11 and 107 pairs of images were selected excluding periods of inactivity or movement. Table 4.4 provides an overview of the selected images per patient.

Tabelle 4.4: Number of selected images for the training dataset.

ID	Selected Images	ID	Selected Images
PA01	50	PA08	38
PA02	50	PA09	96
PA03	60	PA10	101
PA04	81	PA11	99
PA05	99	PA12	38
PA06	107	PA13	58
PA07	11	PA14	71

After selection, image pairs were manually labeled into classes of arms, legs, head, torso and background. Skin covered by clothing and hands or arms of nursing staff were defined as background.

Manual post-processing was conducted on both image modalities due to the low brightness and green color cast in the RGB images, and the low contrast in the LWIR images. The green color cast is caused by the Bayer pattern in the raw image data from the RGB camera. The Bayer pattern interleaves red, green and blue filters on the sensor, with twice as many green sensors as red or blue. Because green is often the most sensitive channel at lower light levels, it's more prone to noise and uneven illumination. This imbalance can result in a green cast in raw images if not corrected by a white balance algorithm. The Gray World Assumption [Buc80] was used for white balance correction, scaling

each color channel (R, G, B) proportionally to achieve a neutral gray average intensity. After correcting the white balance, the images were brightened. If the average pixel intensity across all color channels was less than 70, the image was brightened by a certain factor to enhance darker images more than brighter ones.

Contrast Limited Adaptive Histogram Equalization (CLAHE) was used to improve the contrast in the LWIR images and enhance the visibility of subtle temperature variations. CLAHE is a local histogram equalization technique that adapts to variations in image contrast, preventing the over-amplification of noise often seen with standard histogram equalisation. A pre-processing example for both image modalities can be seen in Figure 4.12.

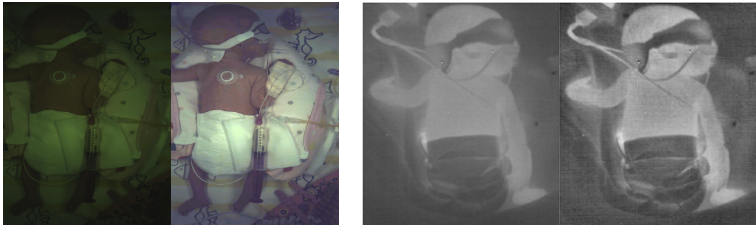


Abbildung 4.12: Examples of unedited and pre-processed images of the infant dataset. Left: RGB image, Right: LWIR image.

To align the LWIR and RGB images, the homography transformation based on the L1 loss algorithm adopted from Section 3.2.1 was compared to a homography transformation using the random sample consensus (RANSAC) algorithm. A reprojection threshold R can be set for the RANSAC algorithm, which causes a pair of points to be treated as an outlier if the absolute error is too large.

For both approaches, six corresponding points on five pairs of images were selected, resulting in 30 corresponding pairs of points. As the camera orientation was slightly different for each recording, the registration was calculated separately for each patient.

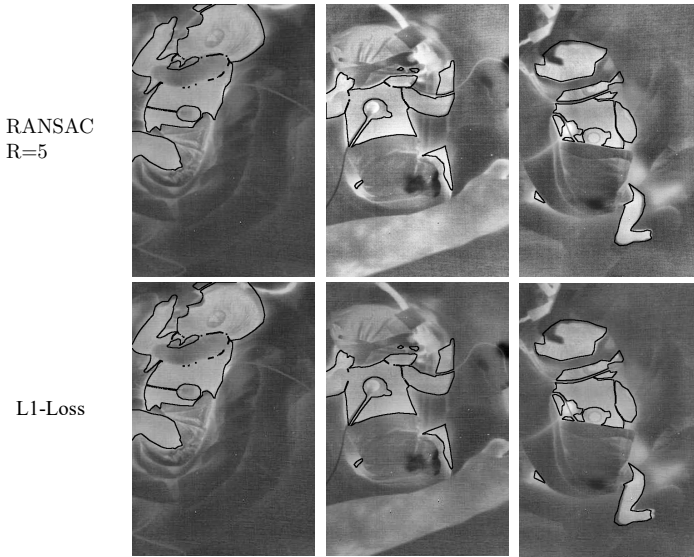
Similar to Section 3.2.1, the corresponding pair points were used to estimate the homography transformation matrices. The RMSE and mean distance were selected as the metrics for evaluating the algorithms for both spatial dimensions. Table 3.2 provides the results of the registration for the L1 loss algorithm and the RANSAC algorithm with $R = 5$, averaged over all recordings.

As in Chapter 3, the L1 loss algorithm produces the best transformation results, with the RANSAC algorithm only slightly inferior. The results were

Tabelle 4.5: Evaluation of the algorithms for registration.

Method	RMSE _x	RMSE _y	mean distance
RANSAC	6.74	6.24	7.06
L1	6.58	4.90	6.82

also evaluated qualitatively, with examples given in Figure 4.13. Here the edges of the masks were transformed to the LWIR images for better visibility.

**Abbildung 4.13:** Comparison of registration approaches; top: registration by the RANSAC algorithm; bottom: registration by the L1 loss approach.

The differences between the two approaches are particularly visible in the left and right images. In the left image, the RANSAC algorithm registers the head more accurately, while in the right image, the RANSAC algorithm registers the leg better than the L1 approach. Due to these superior qualitative results, the RANSAC algorithm was chosen for image transformation.

4.4.2 Segmentation Models

In this chapter, four U-Net-based models have been examined to segment the body parts of infants. The basic architecture of these models, as described in Section 3.2.2, remains consistent: a U-Net structure using a DenseNet-121 backbone for the encoder, with skip connections to enable the transfer of high-resolution information from the encoder to the decoder. To address the specific challenges of the clinical dataset (e.g. variability in image quality, partial occlusions) and to explore the benefits of multimodal data fusion, four different configurations were investigated.

Two models used unimodal image data: one processing RGB images only, the other processing LWIR images only. These unimodal models serve as an important baseline to assess the effectiveness of the fusion strategies. Two additional models integrated multimodal image information through different fusion techniques:

- Simple fusion: This model used two parallel encoders, one for RGB and one for LWIR images. The resulting feature maps from each encoder were then combined by element-wise addition before being fed into the decoder.
- Confidence fusion: Here, two separate U-Nets processed the RGB and LWIR images, producing individual feature maps at their encoders. However, instead of a simple summation, these feature maps were weighted based on a confidence score derived from the output of the respective unimodal models. This reflects the confidence in the classification of each pixel. Higher confidence scores resulted in a greater weight for the corresponding feature maps.

In summary, the four different architectures investigated in this chapter are:

1. Unimodal RGB
2. Unimodal LWIR
3. Simple Fusion (RGB + LWIR)
4. Confidence Fusion (RGB + LWIR)

Furthermore, to overcome the effects of the relatively limited dataset size, the data augmentation methods outlined in Section 3.2.3 were employed for all training. Based on the superior performance demonstrated in Chapter 3, transfer learning was also examined in this chapter, using the CIHP dataset (see Figure 3.6c) for pre-training the model weights.

This multi-level approach allowed for a thorough evaluation of both unimodal and multimodal segmentation strategies, providing a robust analysis of the best strategy for accurate temperature extraction in the context of the clinical data presented in this chapter.

4.4.3 Evaluation Method

Due to the relatively small size of the dataset, cross-validation was used to evaluate the body part segmentation. For this purpose, the dataset was divided into four folds of equal size and an additional test set for final evaluation. The folds contain between 197 and 199 images and consist of two to four recordings. The final validation test set consists of 167 images. All images of an infant were always assigned to a single fold. The split of the dataset is provided in Table 4.6.

Tabelle 4.6: Division of the dataset for cross-validation.

	IDs	Number of images
Fold 1	PA05, PA11	198
Fold 2	PA01, PA02, PA03, PA08	198
Fold 3	PA04, PA06, PA07	199
Fold 4	PA09, PA10	197
Test-Set	PA12, PA13, PA14	167

For training purposes, the categorical cross-entropy was chosen as the loss function. In the confidence fusion training process, the losses from the unimodal components and the decoder were computed and summed. Further, the Adam optimizer with coefficients $\beta_1 = 0.9$ and $\beta_2 = 0.999$ was selected. The networks were trained for a total of 30 epochs with a learning rate of 0.0001.

The segmentation must be highly accurate in order to extract the temperature. This means that the segmentation should only contain correct pixels (true positives) of the corresponding class and minimize incorrectly classified pixels (false positives), such as background pixels. The more correct pixels are extracted by the segmentation, the more robustly the temperature can be subsequently determined. For this reason, IoU and precision were selected as metrics for evaluating the segmentation. The IoU provides information on the size of the segmented region. The precision can be interpreted as a measure of the degree of correctness of the prediction.

4.4.4 Results and Discussion

An initial evaluation of the different model architectures and training configurations was performed using the four-fold cross-validation on the training dataset. The results of this cross-validation, presented separately for unimodal and fusion models, were then used to select the optimal models for subsequent testing on the separate test set.

Cross-Validation

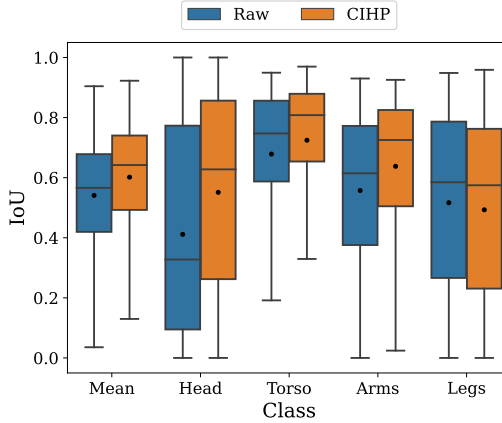
The results of the cross-validation for the unimodal RGB network are illustrated in Figure 4.14. Models without transfer learning (Raw) and with transfer learning (CIHP) were examined.

The model with pre-training achieved the best results with an mIoU of 0.6. In comparison, the model without pre-training achieved an mIoU of 0.5. Overall, the torso was consistently segmented best, followed by the arms and legs. In contrast, the head consistently achieved the lowest IoU, regardless of pre-training. These results are also confirmed by comparing the precision, as given in Figure 4.14b. In this case, the pre-trained model achieves a mean precision of 0.8, while the raw model without pre-training only reaches a mean precision of 0.765. In particular, the segmentation precision of the head is improved by pre-training from 0.65 to 0.76, an increase of almost 0.11. However, the head remains the worst segmented body part in terms of precision. All other body parts achieve good results with a precision of 0.8 or higher. The torso is again the best-segmented body part with a precision of 0.86.

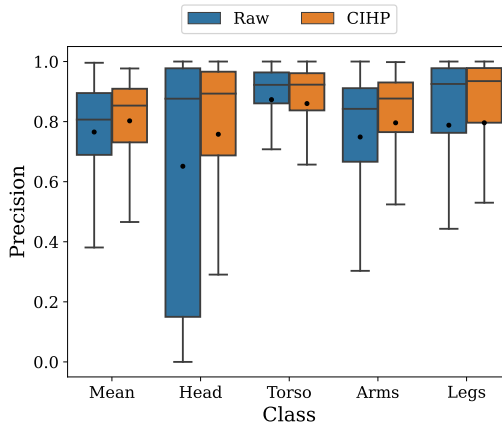
The results of the cross-validation of the LWIR neural network can be seen in Figure 4.15. In addition to the configurations already tested for the RGB images (with and without pre-training), CLAHE was applied to the LWIR images with previous pre-training (CIHP + CLAHE) and investigated.

The configuration of pre-training on the CIHP dataset and applying CLAHE was able to achieve the highest mIoU (0.61) and precision (0.75). The network without pre-training achieved an mIoU of 0.53 and a mean precision of 0.72 and thus achieved the worst results of all LWIR configurations. Overall, the torso was again segmented best and the head worst among all body part classes. However, the segmentation of the legs reaches inferior results compared to the other body parts.

Overall, both unimodal neural networks demonstrated that pre-training significantly improves the segmentation results. This confirms the findings from Chapter 3. Comparing the results of the individual classes, it is evident that the torso class was best segmented. It achieved the best results on the RGB and LWIR models for all configurations. One possible reason for this result



(a) IoU for all classes.

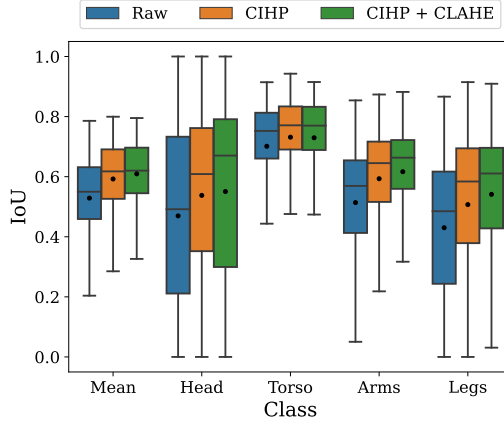


(b) Precision for all classes.

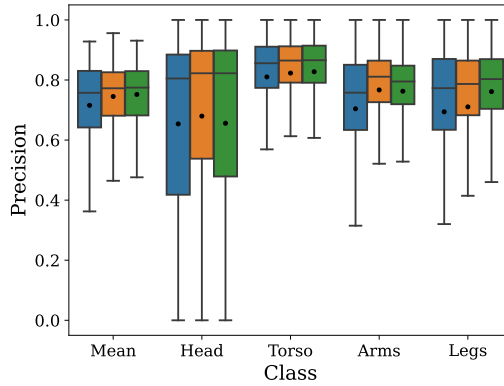
Abbildung 4.14: Results of the cross-validation for the RGB model without (Raw) and with pre-training (CIHP).

is the distribution of the pixels across the different classes. The percentage distribution of the pixels on the training set can be found in Table 4.7.

The cross-validation results described above were generated exclusively on this dataset. Compared to the Chennai dataset from Chapter 3, a larger pro-



(a) IoU for all classes.



(b) Precision for all classes.

Abbildung 4.15: Results of the cross-validation for the LWIR model without (Raw), with pre-training (CIHP) and with pre-training in combination with CLAHE.

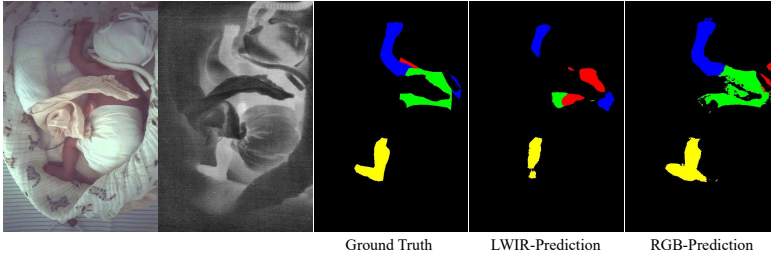
portion of pixels belong to the torso class when the background is discarded. In the Chennai dataset, only 1.8% of all pixels were part of the torso, while in this dataset 5.77% are torso. In addition, the percentage of pixels representing legs is lower than in the Chennai dataset. This could explain the better results of the torso in this chapter, compared to the significantly inferior results in

Tabelle 4.7: Percentage distributions of the classes of the dataset.

	Background	Head	Torso	Arms	Legs
Percentage	83.91 %	4.00 %	5.77 %	3.82 %	2.51 %

Chapter 3.

Additionally, in contrast to the findings of Chapter 3, the head class achieves the lowest IoU in combination with the lowest precision for both unimodal networks. This may indicate that a significant number of pixels are incorrectly assigned to the head class. One potential reason for this could be the presence of head coverings for the infants or respiratory masks, which are challenging to distinguish from the skin due to inadequate lighting or similar temperatures. An example can be seen in Figure 4.16.

**Abbildung 4.16:** Example of the head segmentation by the unimodal networks.

In this image pair, the head is almost completely covered by a hat. In addition, this part of the RGB image is poorly illuminated and the hat appears very bright on the LWIR image. The unimodal networks were both unable to segment the head in this image. This confirms the assumption that head coverings make it difficult to segment the head.

Overall, the LWIR model achieves similar results for both datasets in this chapter and in Chapter 3. However, the RGB results are significantly worse for this chapter. The Chennai dataset was recorded with additional ambient light, while the new dataset was recorded without additional light. This may explain the better results in Chapter 3 due to better contrast/brightness in the images. These findings suggest that using both modalities in the fusion models could improve the segmentation results, as more information about the image is available and dark or blurred images will be segmented effectively.

For the fusion models, both simple fusion and confidence fusion were inves-

tigated. The same training configurations were tested for all models: one with pre-training on the CIHP dataset (CIHP) and another with weights from the best unimodal neural network (UNI). CLAHE was applied to all LWIR input images as it improved segmentation for the unimodal neural networks. The cross-validation results are provided in Figure 4.17.

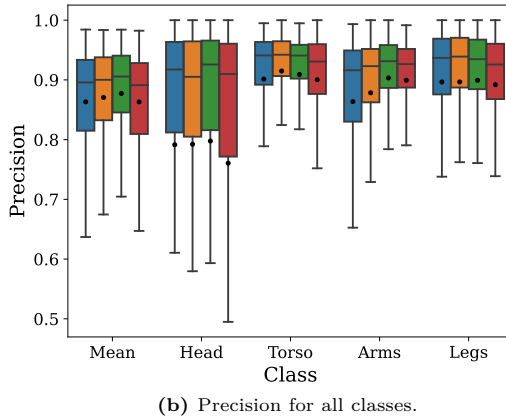
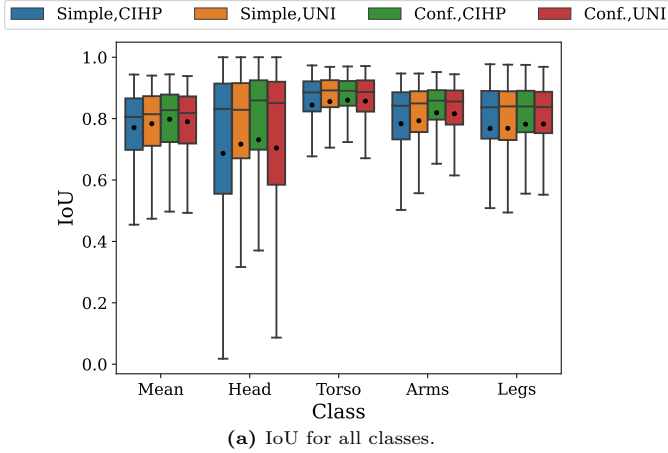


Abbildung 4.17: Results of the cross-validation for the simple and confidence fusion with pre-training (CIHP) and with weights from unimodal models (UNI).

The simple fusion model achieved the best results using the weights of the trained unimodal models with an IoU of 0.78 and a mean precision of 0.87. Only pre-training on the CIHP dataset resulted in a slightly lower mIoU of 0.77 and a mean precision of 0.86. In contrast, the confidence fusion model with pre-training on the CIHP dataset achieved the best results with an mIoU of 0.8 and a mean precision of 0.88. Using the unimodal model weights slightly worsened the segmentation results to an mIoU of 0.79 and a mean precision of 0.86. Similar to the unimodal neural networks, all fusion models performed best for the torso class and worst for the head class.

Evaluation on test set

Following the cross-validation investigations, the best configurations were evaluated on the test set. The four folds of the cross-validation were used as the training set and the models were then validated on the test set. The results are provided in Figure 4.18.

The LWIR model, pre-trained on the CIHP dataset and using CLAHE on the images, achieved an mIoU of 0.389 and a mean precision of 0.62 on the test set. For the unimodal network with RGB images, the model pre-trained on the CIHP dataset was also selected and achieved an mIoU of 0.639 and a precision of 0.727 on the test set. The weights from the best unimodal networks were used for the simple fusion model, which achieved an mIoU of 0.761 and an average precision of 0.812 on the test set. The confidence fusion model, pre-trained on the CIHP dataset only, achieved an mIoU of 0.766 and a mean precision of 0.823 on the test set. Overall, the confidence fusion model produced the best results of all the networks evaluated on the test set, closely followed by simple fusion.

The head remains the worst segmented class across all models, while the torso is consistently the best segmented. Unlike the other classes, the legs are less accurately segmented by the confidence fusion model compared to the simple fusion model. Surprisingly, the unimodal LWIR network performs significantly worse on the test set than in cross-validation. This discrepancy may be due to the division of the training and test sets, as all images of an infant were always fully assigned to a fold or set, possibly making the test and training sets too different. The percentage distribution of pixels from the test set in Table 4.8 also highlights the difference between the test and training sets.

The proportion of background pixels is higher in the test set than in the training set (see Table 4.7), while the proportion of head and leg pixels is lower. In future work, it is important to use a larger dataset and to investigate the effect of dataset partitioning.

In addition, the registration process may have contributed to the poor per-

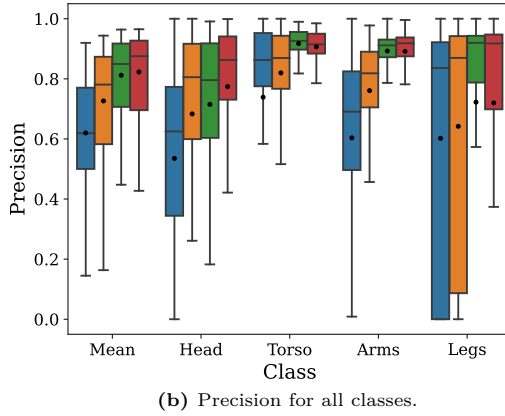
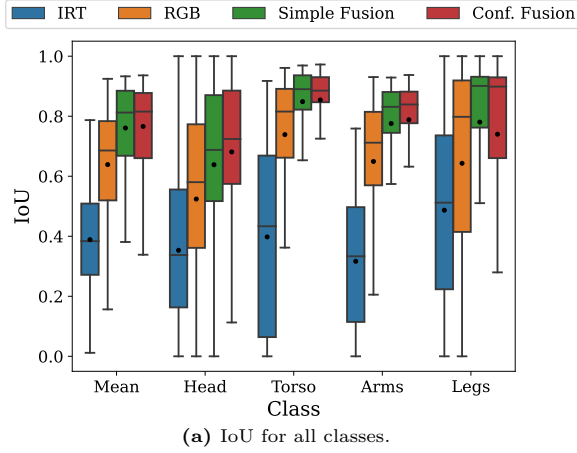


Abbildung 4.18: Results on the test set for the optimized models.

formance of the LWIR model on the test set. It is possible that the images in the test set were registered less accurately than those in the training set. Poor registration could result in LWIR images that are not adequately aligned with the segmentation masks, leading to poorer segmentation results. However, since the RGB images are also available to the fusion models for segmentation, they can compensate for poorly registered LWIR images. For a better under-

Tabelle 4.8: Percentage distributions of the classes in the test set.

	Background	Head	Torso	Arms	Legs
Test set	87.38 %	1.90 %	5.59 %	3.34 %	1.79 %

standing, qualitative examples of the segmentation are provided in Figure 4.19.

Rows three and four give an example of the poor results of the LWIR network. The qualitative predictions reveal that the arms and torso are poorly segmented. This is consistent with the quantitative results in Figure 4.18. One reason for the poor segmentation could be the non-uniformity of these body parts on the LWIR images. The torso and arms have dark and light regions. The dark regions are similarly dark to the background. Comparing the prediction with the LWIR images, it is noticeable that only the lighter regions of the torso and arms were segmented.

All results strongly suggest that the use of RGB and LWIR images in a fusion network improves and stabilizes the segmentation results compared to the unimodal networks. Although the LWIR model gives unsatisfactory results, the combination of both modalities significantly outperforms the RGB model. In particular, confidence fusion was better able to use multimodal images to achieve stable body part segmentation than simply adding the feature.

Despite the overall positive results, there are some limitations. Firstly, the registration of the RGB and LWIR images could be improved. In this chapter, registration was performed by manually selecting image pairs. However, automatic registration is desirable for subsequent automatic temperature measurement. In this context, it would also be possible to develop a smaller and more inconspicuous hardware setup that is installed inside the incubator and is less disruptive during medical care. Another aspect is the size of the dataset, which is relatively small at 959 image pairs. With a larger dataset, more complex architectures, such as transformer-based neural networks, could be investigated for segmentation and possibly lead to even better segmentation. One approach to expanding the dataset could be to generate synthetic training data.

RGB Image	LWIR Image	Ground Truth	LWIR Model	RGB Model	Simple Fusion	Conf. Fusion
			mIoU = 0.58 Prec. = 0.72	mIoU = 0.46 Prec. = 0.71	mIoU = 0.82 Prec. = 0.86	mIoU = 0.82 Prec. = 0.88
			mIoU = 0.33 Prec. = 0.71	mIoU = 0.59 Prec. = 0.80	mIoU = 0.60 Prec. = 0.96	mIoU = 0.67 Prec. = 0.87
			mIoU = 0.09 Prec. = 0.62	mIoU = 0.82 Prec. = 0.92	mIoU = 0.89 Prec. = 0.94	mIoU = 0.89 Prec. = 0.94
			mIoU = 0.47 Prec. = 0.91	mIoU = 0.87 Prec. = 0.90	mIoU = 0.91 Prec. = 0.93	mIoU = 0.92 Prec. = 0.94

Abbildung 4.19: Qualitative results of four sample images from the test set.

4.5 Spatial Temperature Extraction

After examining the segmentation algorithm, this section presents the methods used to extract the temperature of the infants, followed by an evaluation of the data from the clinical trial.

4.5.1 Methods

Since the confidence fusion approach provided the best results, it was used for temperature extraction to segment the body parts. Based on the obtained segmentation masks, the temperature could be extracted. The masks were first transformed in order to align them with the original LWIR images. This was followed by several post-processing steps to increase precision. Next the calibration algorithm from Section 4.3 was applied to the LWIR image. As a last step, different methods for extracting and filtering the temperature from the corrected LWIR image were analyzed. In Figure 4.20 the schematic process of the temperature extraction algorithm is presented.

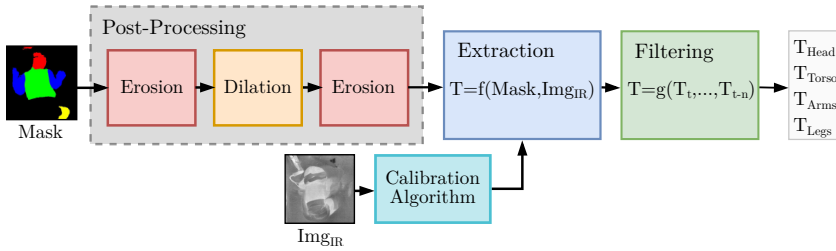


Abbildung 4.20: Temperature extraction process including post-processing, temperature extraction and filtering.

During post-processing, several methods were investigated to increase the precision of the segmentation results. This ensures that only relevant pixels were included in the segmentation masks, excluding pixels from the background or clothing. First, small segmented areas were removed as it is difficult to obtain accurate temperature extraction in these areas. The removal of these small areas consisted of erosion followed by dilation. Separate erosion with different filter sizes was also investigated. This caused a slight shrinkage of the segmentation area to ensure that any misalignment due to the transformation does not lead to the inclusion of background pixels. The post-processing pipeline is visualized in Figure 4.21.

After post-processing, the temperature was extracted based on the pixels of the LWIR image assigned to a class by the segmentation mask. For this purpose, different mathematical methods for the final temperature calculation were investigated and compared with the reference sensor attached to the torso. Averaging and percentile-based thresholding were among the approaches analyzed. For percentile thresholding, the upper and lower percentiles of the temperature values per class were neglected for further extraction. This means

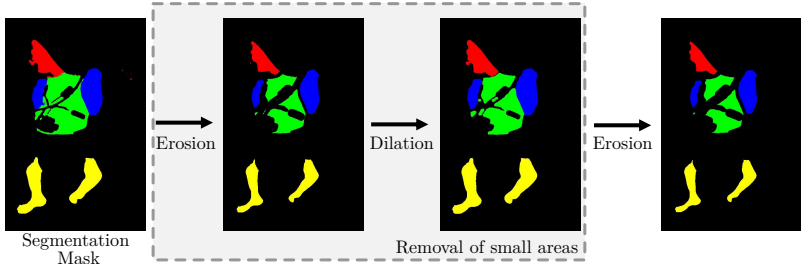


Abbildung 4.21: Example of post-processing: Removal of small regions with subsequent separate erosion.

that the 10% lowest and 10% highest values were not used for temperature extraction in the case of a lower 10/upper 10 percentile. Also, more complex percentiles like lower 50/upper 10 were examined. The final temperature is determined by averaging the remaining values.

In order to obtain a stable temperature measurement, the temperatures obtained after extraction were further refined by applying a Kalman filter. Thermoregulation is a slow, steady process, so no high-frequency changes are expected in the temperature reading. Furthermore, for each of the four temperature regions (head, torso, arms, legs) it was assumed that the temperature is not influenced by any additional known inputs. All temperature variations were modeled as process noise with zero means and assumed variances. This also applies to the observed space. For the process noise Q it was assumed that a change in skin temperature of $0.1\text{ }^{\circ}\text{C}/\text{min}$ is possible. This corresponds to a change of $1\text{ }^{\circ}\text{C}$ per 10 min. This gave a process noise of $Q_{\text{Kalman}} = \frac{0.1^2}{60}$. The standard deviation of the LWIR camera after calibration was used for the measurement noise. This led to a measurement noise of $R_{\text{Kalman}} = 0.12^2$ for the Kalman filter.

Evaluation Metrics

Multiple metrics were defined to provide quantitative information about the temperature extraction. The skin temperature sensor attached to the infant's torso during the recordings was taken as the reference value for determining the metrics. However, this sensor only provides information about the torso.

The mean absolute, median and maximum error were used to assess the accuracy compared to the skin temperature sensor. In addition, the moving standard deviation (movSD) was chosen to assess the robustness of the extrac-

ted temperature itself. Temperature measurement is a non-stationary process, meaning that body temperature can change gradually over time due to factors such as metabolic processes or environmental conditions. The normal standard deviation would not be able to capture these gradual changes and could give misleading information about the stability of the measurements.

For this reason, the averaged movSD with a window size of 200 is used. This metric helps to identify short-term fluctuations in the temperature readings and gives a clearer picture of the consistency and reliability of the temperature extraction process. For example, if the temperature readings remain relatively stable with small variations within each window, this indicates that the extraction method is robust. Conversely, significant variability within a window suggests that the readings may be affected by noise or other inaccuracies.

As the premature infants were admitted over a period of more than five hours, care activities such as diaper changes occurred during this time. During these times, the infants were partially or completely covered, making it impossible to obtain accurate temperature readings. Therefore, to ensure the reliability of the analysis, the metrics were calculated only for the periods without care activities.

4.5.2 Results and Discussion

This section details the results of the temperature extraction process. The confidence fusion approach, which gave the best results, was used to segment the body parts for subsequent temperature extraction. First, the post-processing of the generated segmentation masks was evaluated, followed by an examination of the temperature measurement results. Figure 4.22 illustrates the precision for all classes before and after post-processing.

Overall, the mean precision of the test set was increased from 0.82 to 0.88. The precision of the torso and arms was particularly high after post-processing. The accuracy of the torso improved the most, from 0.91 to 0.97, while the accuracy of the legs improved the least, from 0.72 to 0.76.

With the improved masks, the temperature was extracted according to the pipeline presented in Figure 4.20. The validation was divided into two parts. First, the extracted torso temperature was validated against the recorded reference sensor, to determine the best extraction method (averaging, percentiles). Then the temperature extraction was compared between all body regions.

Since the reference skin sensor was attached to the liver region on the torso, the temperatures of the head, arms and legs are not examined to determine the best extraction method. Both quantitative analyses and qualitative evaluations were employed to assess the accuracy and agreement of the methods.

Figure 4.23 presents a comparative analysis of the methods using box plots.

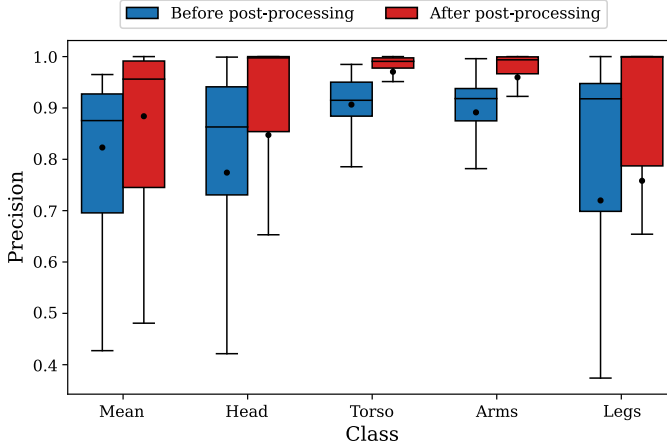


Abbildung 4.22: Precision comparison of all classes before and after post-processing.

This visualization gives a representation of the distribution of absolute temperature errors for four different methods: calculating the mean temperature, disregarding different percentiles including the lower and upper 10th percentile (L10U10), the lower 50th and upper 10th percentile (L50U10) and the lower 65th and upper 5th percentile (L65U05). Other methods were also investigated but resulted in lower accuracies and, thus, were not further examined.

The Mean and Percentile L10U10 methods exhibit similar error distributions, with mean errors of approximately 0.5°C and median errors of 0.51°C . In contrast, the Percentile L50U10 and L65U05 methods, which use higher percentile ranges, demonstrate improved accuracy and precision. These methods yield a lower mean error of 0.44°C and 0.41°C , with a median error of 0.42°C and 0.36°C . The box plots support these findings, revealing a slight downward shift in the median error and a narrower interquartile range for the higher percentile methods. This suggests that using higher percentiles of the segmented pixels not only reduces the mean error but also reduces the variability of the error, suggesting a more reliable estimation of neonatal temperature.

In addition, a qualitative analysis was performed to confirm the feasibility of the L65U05 method, as it provided the best quantitative results. For this purpose, the extracted torso temperature, obtained by disregarding the lower 50th and upper 10th percentile, is compared against the reference temperature in Figure 4.24. For better visibility, only three recordings are presented, with

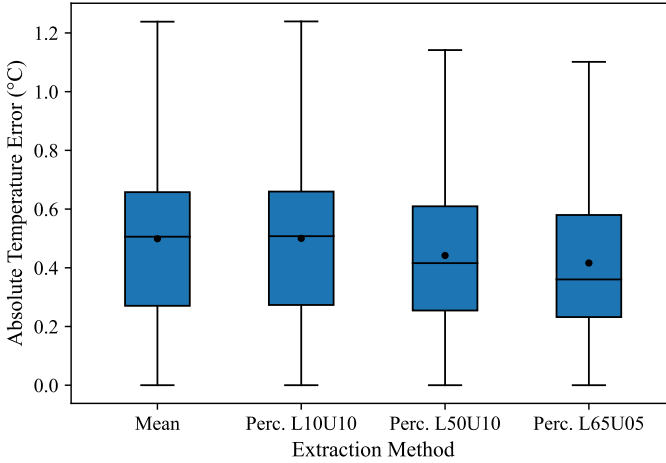


Abbildung 4.23: Comparison of temperature error for different extraction methods.

the gray shaded areas representing the time intervals during which nursing care was provided. In all three examples, the extracted torso temperatures generally follow the trends of the reference temperatures, indicating that the pipeline is capturing the overall thermal dynamics. The bottom plot captures a cooling trend, with both temperature profiles decreasing over the observed period. These consistent patterns of simultaneous heating and cooling further support the ability of the pipeline to detect temperature changes.

Nevertheless, the trunk temperatures often exhibit slight deviations from the reference. In addition, both the reference and the extracted temperatures show short-term fluctuations due to nursing activities. This is particularly visible in the lower plot, where the reference sensor experiences a significant fluctuation around the 170 minute mark. In addition, an unexpected transient drop in the measured torso temperature is observed in the middle plot, which is absent in the other plots.

Overall, the entire temperature extraction pipeline demonstrated a promising level of accuracy for non-invasive neonatal temperature monitoring, with a mean error of approximately 0.4°C . This level of accuracy is encouraging for clinical applications, particularly in scenarios where continuous and uninterrupted monitoring is important. However, there is room for improvement. Although high precision was aimed for during segmentation, subsequent processing reduced the size of the ROI, potentially compromising temperature

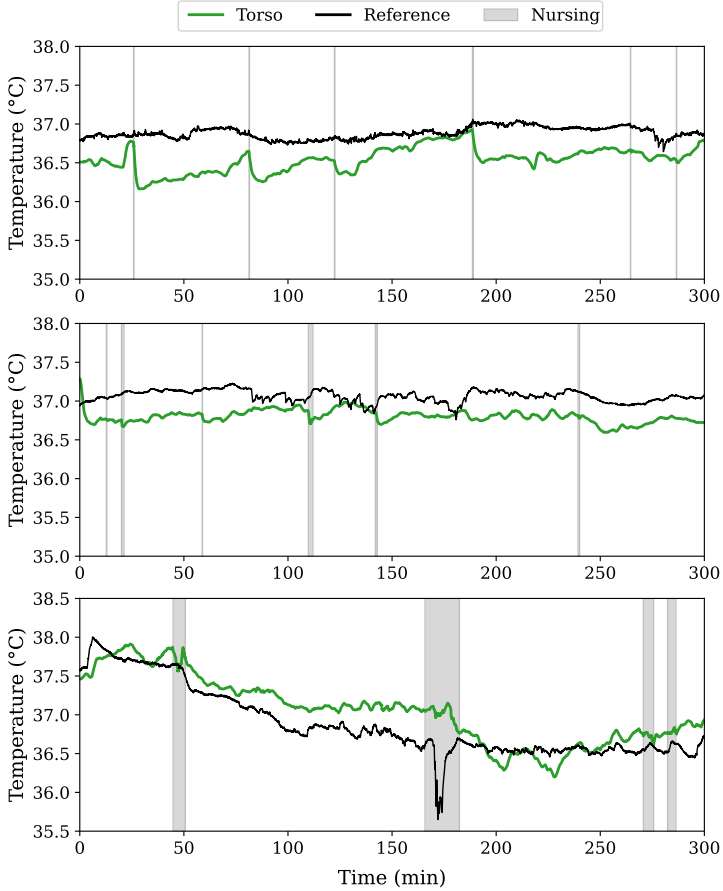


Abbildung 4.24: Extracted temperatures of the torso for three recorded infants compared to adhesive skin sensor as reference.

accuracy. These smaller ROIs may be more susceptible to noise and less representative of the overall body part temperature. Uncompensated thermal radiation from inside the incubator and unknown reflections from the IR window introduce potential systematic errors. Integrating LWIR cameras directly inside the incubator, for example in the upper corners, could eliminate IR window interferences and improve accuracy. The challenge of mapping spatially

derived torso temperature to the single-point reference measurement in the liver region, particularly in prone infants, further complicates the analysis. This raises the question of whether a spatially resolved reference or multiple sensor locations would provide a more suitable ground truth for assessing thermal imaging accuracy. Future work should address these limitations by exploring larger, more robust ROIs, compensating for incubator radiation, and investigating alternative ground truth measurement strategies.

While the torso temperature provides a valuable overall assessment, the accuracy of thermal imaging-based temperature measurements can vary significantly across different body regions due to factors such as skin visibility, motion artifacts and proximity to heat sources. Therefore, the next step was to compare the temperatures obtained from different parts of the body (arms, legs, head) to determine which regions provided the most reliable and robust estimates of central and peripheral temperatures. Both qualitative evaluation and quantitative analysis (movSD) were used to assess the accuracy and robustness of the temperature extraction for each region. Figure 4.25 depicts the extracted temperatures of all body parts, along with the measured temperature of the adhesive skin sensor for reference. Three additional, different recordings are selected, with the gray shaded areas representing the time intervals during which nursing care was provided.

The extracted temperatures for all body parts were within typical skin temperature ranges. However, there was no extracted legs temperature for the bottom plot since the infant's legs were covered throughout the recording. As expected, the torso and head consistently exhibited higher temperatures than the arms and legs, a finding consistent with their proximity to core body temperature. More precisely, the difference in skin temperature between peripheral and central body regions in premature infants is usually between 0.5 °C and 1 °C [Lyo07].

While all body regions had temperature fluctuations, the head and legs demonstrated larger fluctuations than the torso and arms. Furthermore, some of the larger temperature fluctuations observed, particularly in the upper plot of Figure 4.25, can be attributed to interruptions in neonatal care such as diaper changes, examinations and other routine procedures. The qualitative examples suggest that the torso and arms produced the least temperature fluctuation among body parts for the proposed extraction method. This indicates their potential to provide more accurate central and peripheral temperature measurements compared to head and legs when using an LWIR camera.

However, to provide a more complete assessment of the reliability of temperature extraction, both the moving movSD and the percentage of time that each body region was undetectable were examined. Table 4.9 provides the movSD of extracted temperatures, calculated using a 200-second sliding window and

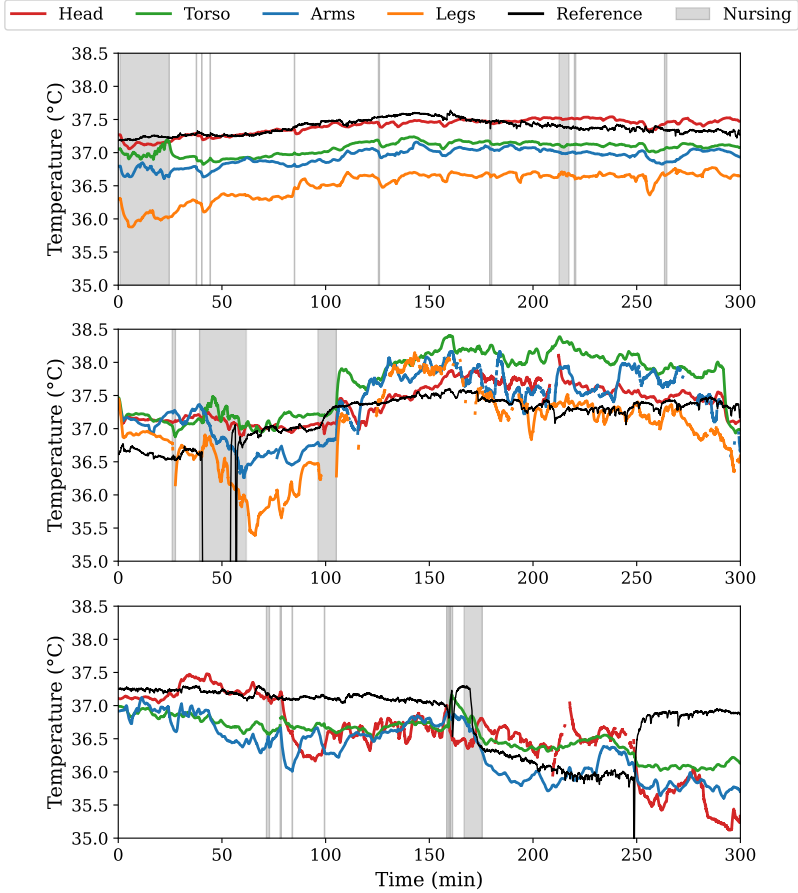


Abbildung 4.25: Extracted temperatures of all body parts for three recorded infants.

averaged over all recorded infants, for each body region (head, torso, arms, legs). In addition, the table includes the percentage of time that each region was not detected during the recordings, providing a measure of the availability of each region for consistent monitoring.

The torso region had the lowest movSD (0.021°C), indicating fewer temperature fluctuations than all other body parts, which have similar movSD. This

Tabelle 4.9: MovSD and non-detection rate of extracted temperature per class, averaged over all recordings.

Metric	Head	Torso	Arms	Legs
movSD (°C)	0.032	0.021	0.032	0.032
Undetectable (%)	4.2 %	0.1 %	1.6 %	10.7 %

supports the initial qualitative observation of greater temperature stability of the torso. Critically, the torso was also the most consistently detected region, with measurements available in 99.9 % of recordings. In contrast, the legs were the least consistently detected, with new temperature measurements unavailable in 10.7 % of recordings, necessitating the use of the last recorded value. The head and arms demonstrated an intermediate detectability at 95.8 % and 98.4 %, respectively. The torso appears to be the most promising region due to its low variability and high detection rate, while the legs are less suitable due to their higher likelihood of occlusion.

Overall, the qualitative evaluation and the quantitative movSD analysis indicate that the torso is the most suitable body region for non-contact temperature monitoring in premature infants using the proposed IRT-based method. The lower movSD for the torso is a sign of superior robustness and accuracy compared to the other regions, especially considering the frequent fluctuations observed due to interruptions by neonatal care. The arms provided reasonably reliable peripheral temperature readings, further strengthening the overall approach for spatial temperature monitoring. The superior performances of the torso and arms align with pixel proportions in Table 4.8, where the torso had the highest percentage, followed by the arms.

Despite the encouraging results, the limitations of this chapter should be taken into account when implementing these findings in clinical practice. Future research is required to validate the proposed methods in larger and more diverse populations and to further refine the IRT-based methodology to minimize the impact of external factors on measurement accuracy.

4.6 Summary

This chapter validates a novel non-contact skin temperature monitoring system for premature infants using infrared thermography (IRT). The system integrated a robust temperature correction algorithm and automated body part segmentation (U-Net fusion model) to enable accurate spatial temperature extraction. This algorithm compensated for systematic errors, including infrared

window transmission effects, reflections from incubator walls and IRT camera drift, and was rigorously validated against a blackbody, demonstrating high accuracy and reliability. The model effectively segmented body parts (head, torso, arms, legs), overcoming the challenges of infant motion and image variability. The complete temperature extraction pipeline demonstrated satisfactory accuracy and robustness compared to a reference sensor.

Despite significant advantages over traditional wired methods, there are limitations. Infant movement artifacts and variations in skin visibility, particularly due to occlusion by clothing or positioning, have a significant impact on measurement accuracy. The relatively small sample size ($N=14$) limits the generalizability of these findings. The complex thermal dynamics within the incubator environment may also introduce further unknown errors. This non-invasive system has significant potential to improve neonatal care. Early detection of thermal stress is critical. Accurate, spatially resolved temperature data could enable early intervention, potentially reducing the incidence of hypothermia. The non-contact nature increases patient comfort, reduces the risk of skin damage and infection, and improves overall well-being. Continuous monitoring facilitates personalized care and reduces the workload associated with manual temperature checks.

Future research should address these limitations. Larger clinical validation studies will improve the generalization. A smaller camera system integrated directly into the incubator would have several important advantages. First, it would minimize interference with clinical care by eliminating the need for external wiring and equipment. Second, it would directly address sources of error associated with the current external camera setup, specifically the effect of the infrared window on temperature readings. Removing this external component would significantly improve the accuracy and reliability of temperature measurements.

5 A Physical Phantom for Neonatal Thermoregulation

Accurate temperature monitoring of premature infants is essential for their optimal management. While infrared thermography offers a promising non-contact alternative to traditional methods, the development and validation of IRT-based systems requires a realistic test platform capable of simulating the complex thermal behavior of premature infants. This chapter presents a novel neonatal phantom that addresses this need. The phantom incorporated a 3D-printed anatomical structure equipped with thirteen independently controlled heating zones and embedded Pt1000 temperature sensors to realistically mimic the thermal dynamics of a premature infant. A silicone skin layer enhanced with boron nitride for improved heat distribution enveloped the phantom, providing a lifelike appearance and tactile experience. The phantom's ability to reproduce both physiological and pathological temperature scenarios, such as an increased central-peripheral temperature difference, was validated through a series of tests. These results demonstrate the effectiveness of the phantom as a test platform for the development and validation of non-contact temperature measurement systems and control algorithms for neonatal incubators. Parts of this chapter have already been published in peer-reviewed journals and conferences [VLB⁺ 22], [VTL⁺ 23].

5.1 Motivation and Overview

As discussed in the previous chapter, non-invasive measurement methods based on infrared thermography are a current subject of research for neonatal care. However, the creation of such methods requires large amounts of (image) data [VDDP18]. The challenge lies in the limited availability of neonatal image data. In this context, thermal phantoms can be used to develop contactless temperature measurement systems and, additionally, investigate the thermal behavior of premature infants.

Thermal phantoms of premature infants were developed in various works. In 1992, Sarman et al. [SBHT92] constructed a 1 kg phantom to examine non-evaporative heat loss, providing a method for assessing thermal conditions in neonatal care. Furthermore, in 1999 Apedoh et al. developed a thermal phantom that replicated a full-term infant to determine the contributions of different components of dry heat exchange [AEHT⁺99]. Elabbassi et al. created a phantom and investigated the dry heat losses of small-for-gestational-age newborns to compare differences in the ability to lose heat between prone and supine positions [EBM⁺01]. A similar approach was taken a few years later by Belghazi et al. with a phantom also made of copper. This phantom was able to simulate an evaporated heat emission by distributing water on the surface of the phantom [BTE⁺06]. Further, in 2016 Ostrowski et al. [ORS⁺16] performed

a study to assess the body dry heat loss from an infant in a radiant warmer using a copper cast anthropomorphic thermal phantom.

In recent years, Kurcinska et al. [KSK19] developed a thermal phantom specifically designed to measure the insulating power of garments. In 2021, Han-nouch [Han21] built a 3D-printed thermal phantom to resemble a premature infant and studied the heat transfer from the phantom under various conditions.

Recently, Lyra et al. [LVC⁺21] developed a phantom that can heat up to a desired temperature and simulate photoplethysmographie (PPG) signals. The size corresponded to a premature infant of the 32nd week of pregnancy and the outer skin was made of skin-colored silicone. However, this phantom also had some limitations. Since the heating elements were only attached to the front, the back could not be heated. In addition, the neonatal phantom lacked an outer layer such as black acrylic paint, which means its emissivity was not as high as that of human or infant skin. This is a problem because accurate emissivity is critical to replicating the thermal characteristics of infant skin, which in turn is essential for reliable thermal imaging. Furthermore, different body positions could not be simulated, since no joints have been integrated into that phantom.

In this chapter, a novel neonatal phantom is introduced, capable of simulating the thermoregulation of premature infants in both physiological and pathological temperature scenarios. The outer shape of the phantom is adapted to the body of premature infants. With the help of resistive heating elements, the skin temperature can be regulated separately for different areas of the body. Additionally, the phantom features a layer of black acrylic paint with high emissivity, making it viable for thermal imaging. First, in Section 5.2.1, the systematic procedures for the development of the phantom will be explained. The 3D model and the modification steps carried out are discussed. Section 5.2.2 will describe the operation of the phantom with its components. Afterwards, the control system including the printed circuit board will be introduced in Section 5.2.3. In Section 5.3, the results of the thermal homogeneity tests and the validation of the thermal phantom will be presented and discussed. Finally, in Section 5.4 the most important findings will be summarized.

5.2 Materials and Methods

5.2.1 3D-Modeling

A model of an infant of indeterminate age [Kim19] was modified to simulate the physiological characteristics of a premature infant. The infant model had a lower limit for scaling due to the spatial constraints for the internal components

and elements. The torso, which was designated to house the circuit board, was a critical factor in this regard. Consequently, the model was scaled to a length of 38 cm and weight of 1250 g, which corresponds to the median height and weight of premature infants at the 29th week of gestation [FK13]. Additionally, the total skin surface of 0.108 m^2 corresponds to a premature infant at 28th week of gestation [Ahn10]. The model was then segmented into six distinct body segments (head, torso, and upper and lower limbs). Moreover, a scaffold was created for each body segment, where heating elements and sensors can be attached to. Subsequently, the scaffolds for all body segments were partitioned into an upper and a lower half, facilitating better cable management and 3D printing. All modifications of the 3D model were performed in *Autodesk Fusion 360* (Autodesk, San Rafael, USA). The altered model of the infant (a) and its scaffold (b) is depicted in Figure 5.1.

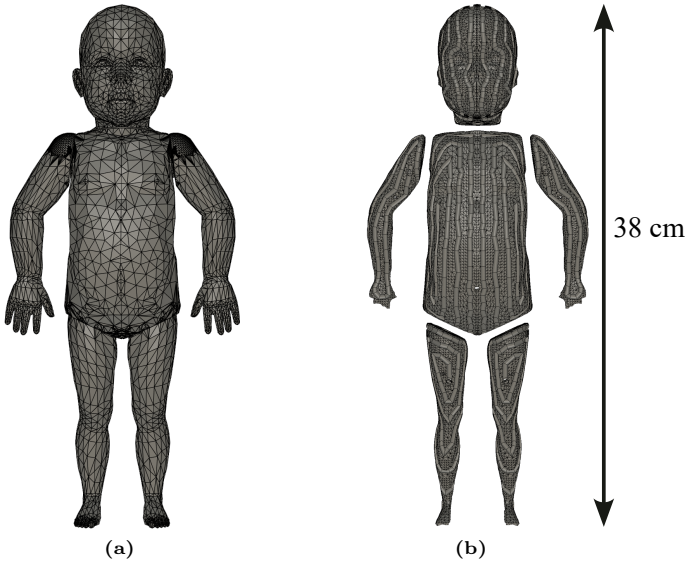


Abbildung 5.1: a) Modified 3D model of infant; b) Scaffolds for all body parts.

A hatch was incorporated into the hollow torso, enabling the integration of a printed circuit board (PCB) with the power electronics. The cables can be routed through a 10 mm diameter hole. The shape of the joints allows the body parts to be tilted by a few degrees and additionally rotated completely. The

placement of the joints between the body parts means that they are not in direct contact with the torso.

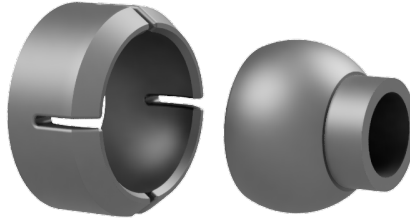


Abbildung 5.2: Ball joint with 10 mm hole for the cables, consisting of ball socket (left) and joint head (right)

The ball socket (Figure 5.2 left) is attached to the torso and the joint head (Figure 5.2 right) to the corresponding body part. The diameter of the joint head is 15 mm with a 10 mm diameter hole for routing the cables.

5.2.2 Thermoregulation

The phantom was designed to control the temperature in 13 distinct areas individually. An overview of the 13 heating areas in a thermal image is provided in Figure 5.3.

The arms, legs and head each consist of two heating areas. For these body parts, the temperature of the upper and lower parts (corresponding to the frontal plane) can be adjusted separately. The torso is divided into an upper section and two lower sections, with the hatch representing a separate heating area. Carbon fibers of type *Kohleroving Tenax HTS40 F13 / 12K / 800 tex* (R&G Faserverbundwerkstoffe GmbH, Waldenbuch, Germany) were used as heating elements. The strands are a composite of 12000 uncoated carbon fibers. The individual fibers (filaments) have a diameter of 7 μm . The starting material for manufacturing the fibers is the polymer polyacrylonitrile.

The carbon fiber strands represent a low ohmic resistor with a resistance per length of $R' \approx 34 \Omega/\text{m}$. Applying an electrical voltage leads to a power dissipation to the fiber strand, which is completely converted into heat. Depending on the size of the heating area, the length of the fiber strands per heating element varies. For the underside of the arm, it is only 36 cm. In contrast, the top of the torso consists of 230 cm fiber strands, which results in a significantly higher resistance value.

The heating areas need to be accurately measured to ensure stable temperature control. For this purpose, Pt1000 temperature sensors from the manufac-

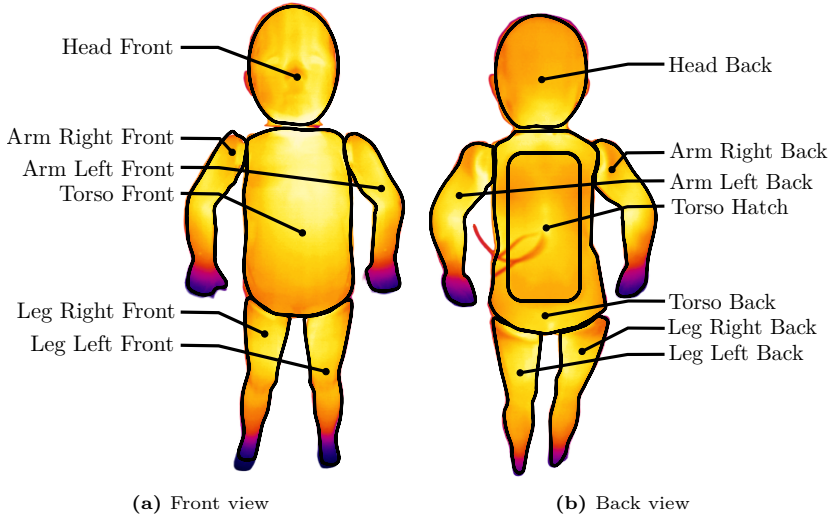


Abbildung 5.3: Heating areas of the thermal phantom

turer Heraeus Nexensos were used. These sensors are known for a high accuracy of $0.1\text{ }^{\circ}\text{C}$ and compact design. Additional specifications can be found in Table 5.1.

Tabelle 5.1: Specifications of the temperature sensor Pt1000 from the manufacturer Heraeus Nexenso.

Property	Value
Type	M 422 A PT 1000
Size (L×W×H)	4 mm × 2.1 mm × 0.9 mm
Temperature range	-50 °C to +300 °C
Tolerance class for 0 °C to +150 °C	Class F 0.1 (1/3 B)
Temperature coefficient	3850 ppm/K

Finally, the temperature sensors and carbon fibers were placed on top of the previously created scaffolds. This allows the temperature to be accurately measured at the skin surface. In addition, the fibers were distributed as evenly as

possible to achieve a homogeneous heat distribution. A fully assembled scaffold is illustrated in Figure 5.4 for the example of a left arm.

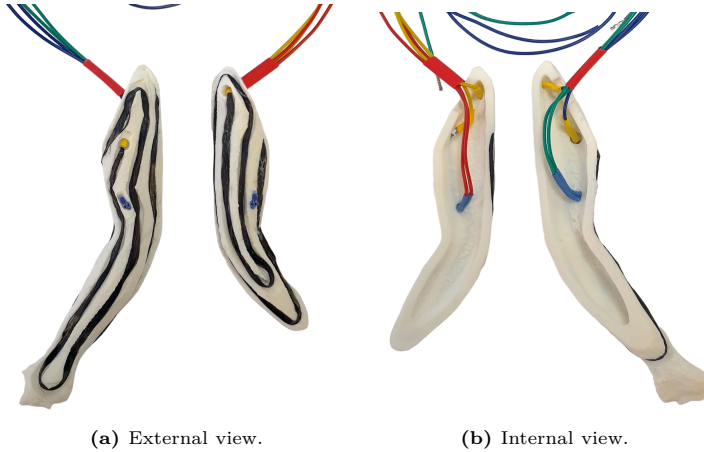


Abbildung 5.4: Equipped scaffold of the (left) arm with heating and sensor elements.

Each scaffold of the body parts was finally encased in a layer of silicone. The two-component silicone SILIXON10 was used for this purpose, which was then poured into a 3D-printed mold around the corresponding body part. Heat is exchanged between the carbon fibers and the silicone layer: The warmer heating elements transfer thermal energy to their colder surroundings – the silicone layer. Conduction is the dominant type of heat transfer here. The heated silicone layer exchanges heat with the surrounding air via convection. When the phantom is in direct contact with a surface, such as a mattress, additional conductive energy exchange takes place at the contact surface. The silicone layer also acts as a protective layer that can protect the fiber strands from damage.

5.2.3 Thermal Control

A PCB was designed to control the heating elements and read the sensor values. The circuit board allows the connection of one temperature sensor and one heating element each for 14 channels. The board was designed and manufactured by *Dieter Richrath Elektronik GmbH*. A microcontroller of the type

PIC32MX795F512L from the manufacturer *Microchip*) was used, which allows serial communication via a Universal Asynchronous Receiver/Transmitter interface. The board has a compact design. This allows it to be integrated into the phantom through the phantom hatch as depicted in Figure 5.5.

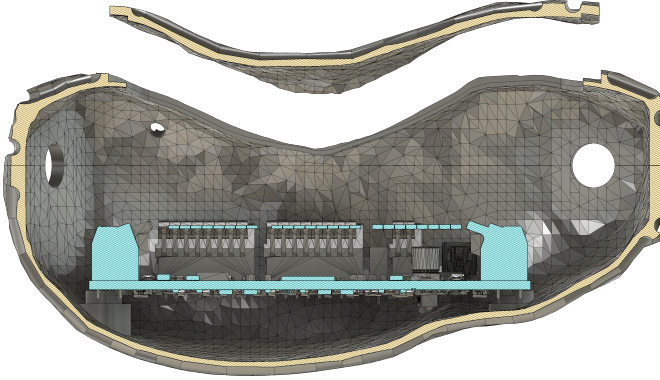


Abbildung 5.5: Sectional view of torso with the circuit board.

The board allows accurate reading of Pt1000 resistors using sigma-delta analog-to-digital converters (ADCs). An increase in temperature leads to a measurably lower conductivity of the sensor. The accuracy of the measurement is given as 0.1°C . The power conversion in the heating elements is controlled by pulse width modulation (PWM). Figure 5.6 illustrates the schematic of the heating elements for one channel. The heating elements are connected in series with the power transistors on the circuit board. The gate of the transistors is connected to the microcontroller via a $10\text{ k}\Omega$ resistor. When turned on, the transistors are saturated, therefore, most of the voltage drops across the heating elements. When turned off, the entire operating voltage drops across the transistor.

When the board is connected to a computer via a USB cable, certain commands can be sent to the microcontroller via serial communication and information about the temperature status can be received. The values are output every second, providing the temperature in degrees Celsius to the first decimal place. The PWM value can be set for each channel with an accuracy of 0.1% . The board does not have its own heating power control. However, serial communication can be used for external heat control.

A schematic overview of the phantom's communication and functionality is given in Figure 5.7.

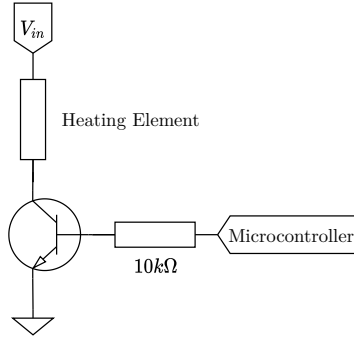


Abbildung 5.6: Circuit diagram for the heating areas of the thermal phantom

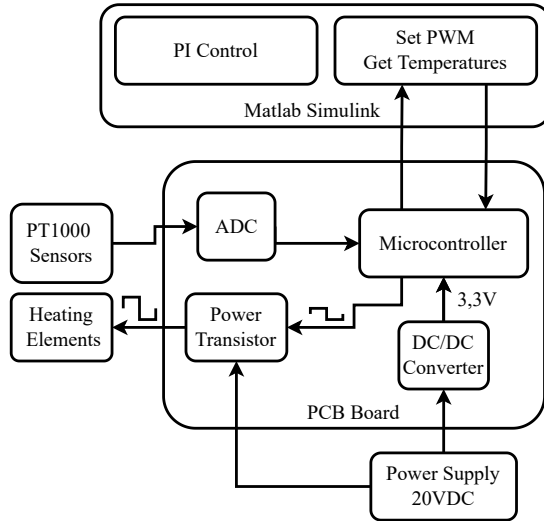


Abbildung 5.7: Schematic functionality and communication of the phantom.

Finally, the controller was implemented using Matlab and Simulink software (The MathWorks, U.S.A.). A proportional integral (PI) controller was designed for each heating zone based on a recorded step response.

5.2.4 Optimization of Thermal Homogeneity

Due to the non-uniform heat dissipation of the integrated carbon fibers, there is a risk of inhomogeneous heat distribution on the phantom surface. To compensate for this, a high thermal conductivity of the outer layer is advantageous.

The thermal conductivity of the SILIXON10 silicone used is not known. However, it is generally around 0.2 W/m/K for silicone compounds [Int18]. Various materials can be combined with silicone to improve conductive heat transfer. Graphite or boron nitride have already been used for this purpose in several experiments [KOA10, ZWZ⁺21, YZZ⁺18]. When selecting the material, it is also important to ensure that the mixed silicone mass has the lowest possible electrical conductivity, otherwise a current may flow in the silicone skin. According to the Wiedemann-Franz law, there is a proportionality between thermal and electrical conductivity for many electrically conductive materials [HMS21].

Graphite is a very good electrical and thermal conductor. Adding it to silicone can also increase the electrical conductivity [KOA10]. boron nitride, on the other hand, is a very good thermal conductor, although it has a low electrical conductivity [Hen24]. Kemaloglu et al. [KOA10] introduced boron nitride particles of different sizes into the matrix of silicone. It was found that the thermal conductivity of the silicone-boron nitride mixture increases with the boron nitride content. The thermal conductivity coefficient was increased to 2 W/m/K for a maximum content of 50 %. In addition, particularly large, plate-shaped boron nitride particles allowed for higher thermal conductivity.

Thermal Simulation

Based on the results presented by Kemaloglu et al. [KOA10], a thermal simulation was performed in Fusion 360 to investigate the effects of an increased thermal conductivity coefficient in the silicone material. Instead of simulating the entire newborn phantom, a simplified model of one arm was created, as the computational effort and simulation time increased significantly with growing geometry complexity. The simplified model is depicted in Figure 5.8.

Materials have been assigned to the modeled solids. For the silicone skin, the material rubber-silicone from the library was used. The thermal load *internal heat* with a power of 2 W was assigned to the heating elements. In addition, the silicone surface was assigned the loads *convection* and *radiation*. For these loads, an ambient temperature of 25°C , an emissivity of $\epsilon = 0.9$, and a convective heat transfer coefficient of $h_c = 5 \text{ W/m}^2/\text{K}$ were set. The steady-state thermal behavior was studied. The simulation was performed for three different scenarios:

1. Material 1: Pure silicone ($K_c = 0.23 \text{ W/m/K}$),

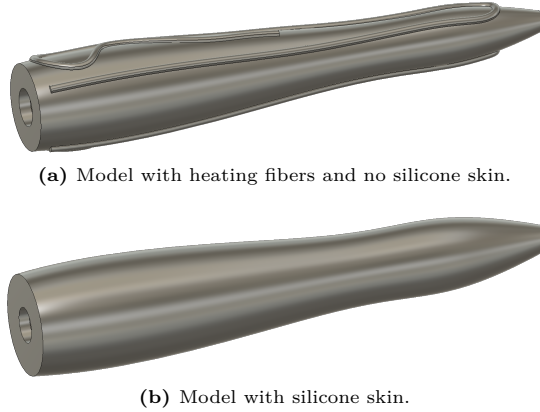


Abbildung 5.8: Simplified model of an arm for thermal simulation.

2. Material 2: A boron nitride-silicone blend ($K_c = 0.75 \text{ W/m/K}$), corresponding to a filling ratio of 10-30 %, depending on the boron nitride.
3. Material 3: High thermal conductivity coefficient of $K_c = 2.3 \text{ W/m/K}$ for reference

The simulation results are provided in Figure 5.9 and include the temperature distributions on the surfaces, visualized using a rainbow color scheme.

As indicated in Figure 5.9a, the warmer areas for material 1 appeared on the upper side of the model, with a maximum temperature of 46.3°C . A clear decrease in temperature was noticeable towards the right side of the model, which resulted in a temperature difference of 18.9°C between the maximum and minimum temperatures. For materials 2 and 3, the heat distribution demonstrated less variation. The temperature for material 2 ranged from 40.0°C to 40.7°C , while for material 3, it ranged from 39.1°C to 41.0°C .

In summary, materials 2 and 3 exhibited a significantly more homogeneous heat distribution compared to material 1. Interestingly, although material 3 had a higher thermal conductivity than material 2, its temperature difference was slightly greater.

Evaluation with physical models

Based on the additional simulation results, boron nitride was chosen as the filling material for the silicone. The product HeBoFill CL-SP 035 from the

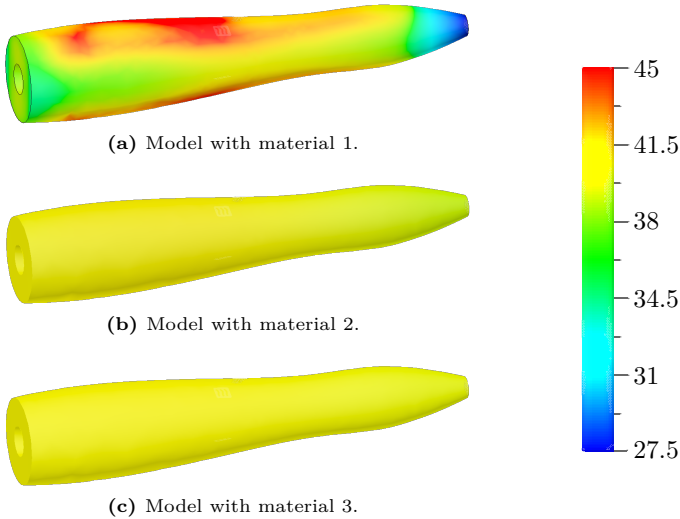


Abbildung 5.9: Temperature distribution on the surface of the simulated models.

manufacturer Henze boron nitride Products AG (Lauben, Germany) was employed. This boron nitride powder is 99 % pure and has a coarse grain structure. It boasts good thermal conductivity and high purity, allowing for high filling levels without significantly increasing viscosity. The average particle size is approximately 35 μm .

The thermal simulation has provided evidence that increased thermal conductivity allows for better heat distribution on the surface. Therefore, the highest possible boron nitride content in the silicone appeared desirable. However, this could not be amplified indefinitely, as the viscosity increases with the amount of powder added. Therefore, a series of tests was conducted to evaluate the viscosity at different mixing ratios. The goal was to determine the maximum amount of boron nitride that would still allow the molds to be filled. The maximum viscosity was found to be reached at a boron nitride content of about 16.7 percent. Before manufacturing all the body parts with the boron nitride silicone, different test models were produced (see Figure 5.10), to determine the best boron nitride ratio for optimal thermal homogeneity.

As a result, arm models were made with 16.7 % (Figure 5.10 left) and with 7 % (Figure 5.10 center) boron nitride. For comparison, an arm model with pure silicone skin (Figure 5.10 right) was also produced. The only difference between

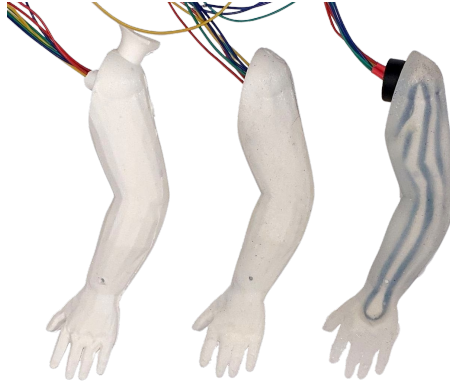
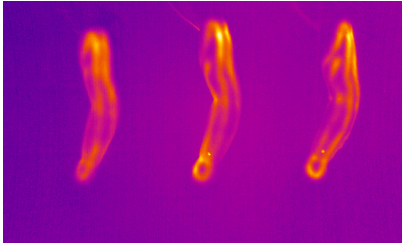


Abbildung 5.10: Different boron nitride contents in the silicone for the test models: left 16.7 %, middle 7 %, right 0 %

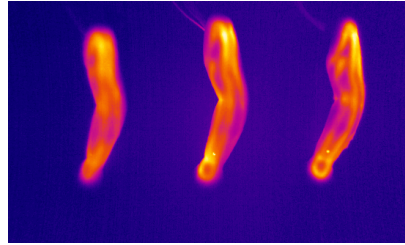
all test models was the silicone skin. Afterwards, the thermal behavior of the three arm models was investigated and compared in a heat-up test. The same heat output of 4 W was generated in all physical models. Three separate power supplies were used and turned on simultaneously. The heating process was observed with a PI 640 thermal imaging camera from Optris GmbH. Initially, the temperature was uniform across the arms, with heating beginning after the power source was switched on.

The thermal images recorded during this process are presented in Figure 5.11. After 30 seconds (Figure 5.11a), the position of the heating elements in the silicone can be clearly seen, as the silicone in the immediate proximity of the heating elements heats up first. Warmer areas could also be seen on the upper side (corresponding to the shoulder area of the arm). Another 30 seconds later (Figure 5.11b), the temperature distribution became more homogeneous, and the heating elements were less visible in the thermal image. The greater the concentration of boron nitride in the silicone, the more evenly the temperature was distributed throughout the arm. This trend continued in the following thermal images (Figure 5.11c and 5.11d), where an increasingly homogeneous temperature field formed on the surfaces of the body parts.

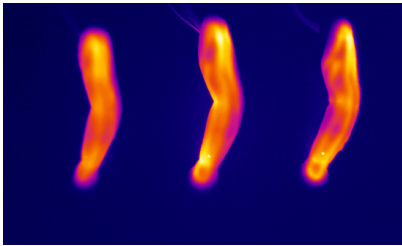
After six minutes (Figure 5.11e), the heating elements in the forearm area of the body part without boron nitride were still clearly visible. In the arm with 7 % boron nitride, these were no longer visible, but the temperature in the area of the crook of the arm increased. The arm with 16.7 % boron nitride formed the temperature field with the most even distribution. For all three arm bodies,



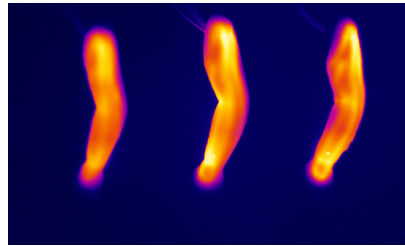
(a) After 30 sec, maximum temperature 27 °C.



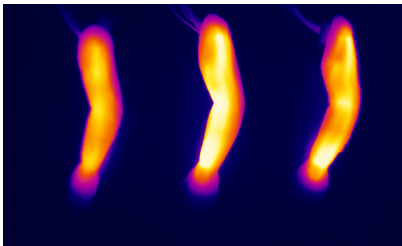
(b) After 60 sec, maximum temperature 31 °C.



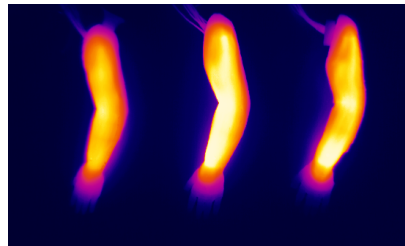
(c) After 90 sec, maximum temperature 34 °C.



(d) After 120 sec, maximum temperature 38 °C.



(e) After 6 min, maximum temperature 53 °C.



(f) After 12 min, maximum temperature 62 °C.

Abbildung 5.11: Temperature distribution in the heating process. Left 16.7%, in the middle 7%, right 0% boron nitride content in the silicone. The color scaling is adapted to the temperature range per image.

a warmer area could be seen at the wrist, which was more prominent with low boron nitride contents. The thermal images after twelve minutes (Figure 5.11f)

hardly differed from those after six minutes except for the uniform increase in temperature of all test arms. The body parts with a high boron nitride content demonstrated a uniform thermal image during the experiment. For these reasons, a silicone mixture with a boron nitride content of 16.7 % was selected for the construction of the neonatal phantom. Besides optimizing temperature uniformity, a layer of black acrylic paint was applied to the silicone surface to achieve an emissivity of $\epsilon = 0.98$, similar to human skin.

5.3 Results and Discussion

The final neonatal phantom is depicted in Figure 5.12 together with a corresponding thermal image, which was taken after a long heating period in a stationary state.

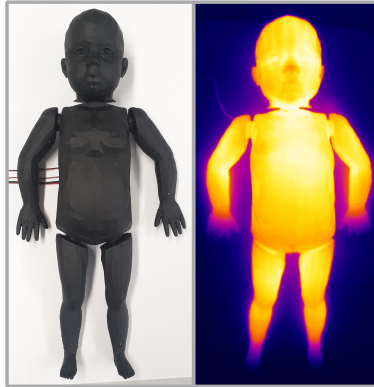


Abbildung 5.12: The final neonatal phantom (left) and a thermal image recorded with an infrared camera (right).

The assembled phantom has the shape and external appearance of a premature infant. In addition, it has a homogeneous temperature distribution due to the added boron nitride. Only the hands and feet are noticeably cooler, as no heating elements could be integrated there, especially in the fingers and toes.

Furthermore, Figure 5.13 provides thermal images with different body poses of the neonatal phantom. As the arms, legs and head are attached to the torso with ball joints, these body parts can be placed in various positions. The legs can be moved at an angle of up to 45° to the sagittal plane. It is also possible to rotate the arms, legs and head completely around the joint axis. The arms can therefore be placed against the torso or stretched over the head.

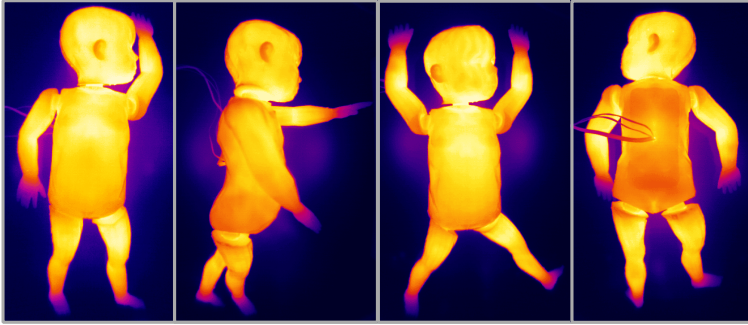


Abbildung 5.13: Thermal images of the phantom in various poses.

The thermal images, presented in Figure 5.13, also confirm the uniform temperature profile for other body poses, especially for a prone position. However, the cables that are necessary for the control and power supply can be seen here.

After the qualitative evaluation, the phantom was validated quantitatively in three steps. First, the temperature control of the heating areas was validated using the internal Pt1000 sensors. Then a surface-based skin temperature control was evaluated and finally, pathologies were simulated.

5.3.1 Validation of temperature control

To validate the temperature control, all heating areas were set to 37°C . The temperature values from the internal Pt1000 sensors were then recorded over a period of 33 minutes. Figure 5.14 illustrates the temperature validation for only four heating areas, providing a better visualization.

During the time interval from 0 to 300 seconds, a steep, linear temperature rise is observed across all heating regions. During this initial time interval, the temperature rise is primarily dependent on the maximum achievable power conversion within the heating areas, which is limited by an area-dependent maximum PWM value. Then, between 300 and 500 seconds, the curves begin to flatten out. The temperatures gradually approach the setpoint and eventually oscillate around it with a maximum deviation of $\pm 0.4^{\circ}\text{C}$.

In a further test, the temperature behavior of the phantom was examined when the setpoint was changed from 36.5°C to 39°C . The phantom's controller was in a steady state at the time of the change. The corresponding temperature responses are depicted in Figure 5.15 for only four heating areas, providing a better visualization.

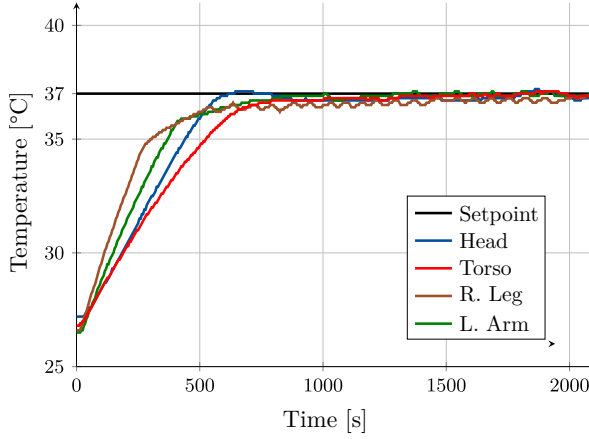


Abbildung 5.14: Temperature control of the heating areas to a setpoint of 37°C, measured with internal Pt1000 sensors.

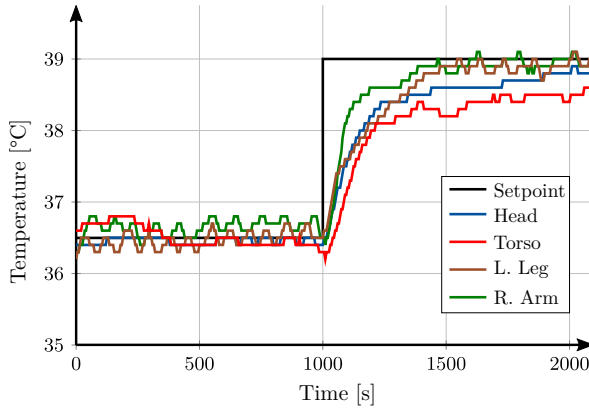


Abbildung 5.15: Temperature control of the heating areas for a setpoint change from 36.5°C to 39°C at 1000 seconds, measured with internal Pt1000 sensors.

The setpoint was changed after 1000 seconds. A rapid, linear increase in temperature was then observed in all regions for a short period of time. However, the time required to return to a steady state was different for each controller. While the arms and legs needed about 500 seconds to stabilize, the torso

controller did not reach this state until 1000 seconds after the setpoint change.

The validation of the neonatal phantom confirmed a stable control behavior based on the internal Pt1000 sensors, with a slight oscillation present. Furthermore, the time required to reach a steady state varies widely within the heating zones.

5.3.2 Skin Temperature Control

As the Pt1000 sensors are embedded in the silicone, they do not correspond to the surface temperature. The internal sensors therefore are only point sensors in the heating areas. However, surface-based temperature control would be beneficial for simulating pathological temperature scenarios. For this purpose, the surface of the heating areas was recorded with the same calibrated infrared camera as before. The difference between the temperature of the internal Pt1000 sensors and the temperature measured by the infrared camera is given in Figure 5.16 for a 37°C heating process. A variable temperature difference

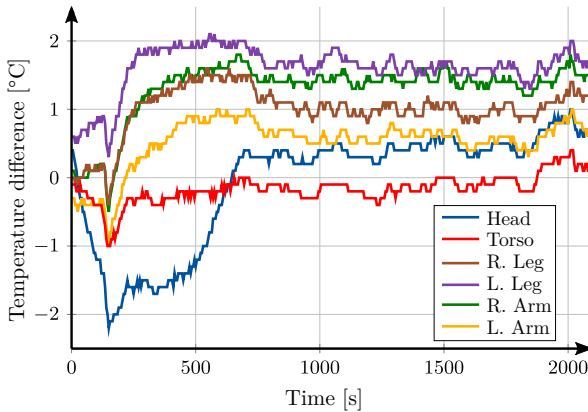


Abbildung 5.16: Temperature differences of the heating areas measured by the internal sensors and the external infrared camera.

can be observed at the beginning of the recording. However, after 1000 seconds a steady state is reached for the temperature differences of all heating areas. Thus, a steady-state offset can be determined and used to set up a control system based on surface temperatures. When designing surface temperature control using internal sensors, it is important to consider this offset by adding it to the set temperature for each heating zone. The offset values are given in

Table 5.2 for the front heating areas.

Tabelle 5.2: Temperature offsets between skin temperature and internal Pt1000 sensors during the steady state of frontal heating areas.

Heating Area	Offset
Head	0.5 °C
Torso	0 °C
Arm Left	0.8 °C
Arm Right	1.5 °C
Leg Left	1.7 °C
Leg Right	1.3 °C

The offset values from Table 5.2 were used to control the average skin temperature to 37 °C. The averaged temperatures of the regions of interest, measured with the infrared camera, are presented in Figure 5.17 for only four heating areas, providing a better visualization.

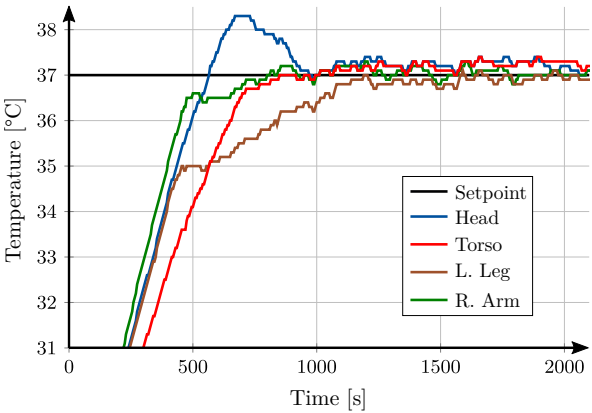


Abbildung 5.17: Temperature for offset-adjusted control to a setpoint of 37 °C, measured by the external infrared camera.

An overshoot of the head area was observed. However, all areas reached the target value of 37 °C surface temperature for a time greater than 2000 seconds. At this steady state, the maximum deviation of the temperature from the

setpoint was $\pm 0.4^\circ\text{C}$ in all areas.

This section has demonstrated that it is possible to ensure stable control of skin temperature, observed by an external infrared camera. Hence, this approach allows for a uniform temperature distribution throughout the thermal phantom.

5.3.3 Pathology Simulation

Finally, a pathological scenario was simulated using the developed surface temperature control of the neonatal phantom. For this purpose, the scenario of an increased cpTD, which is an early sign of late-onset sepsis in premature infants, was simulated. The target temperatures of the phantom's arms and legs were set at 36°C and those of the torso and head at 38°C , resulting in a cpTD of 2°C . The averaged temperatures recorded with the infrared camera are provided in Figure 5.18.

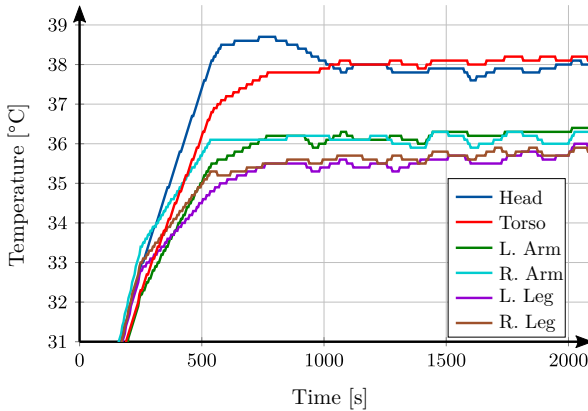


Abbildung 5.18: Temperature profile for centralization with a temperature difference of 2°C , measured with the IRT camera.

Starting at $t = 300$ seconds, it is evident that the temperatures of the different body parts approached different final values. After 1000 seconds, the head, torso, arms and legs have reached the target value, demonstrating the simulation of an increased cpTD.

The corresponding infrared images are illustrated in Figure 5.19. On the left, the physiological temperature distribution of a healthy infant is depicted. On the right, a central-peripheral temperature difference of 2°C has been simula-

ted. The different warm areas can also be seen in the false color visualization, as the torso and head are lighter than the arms and legs.

Overall, the phantom is capable of simulating both physiological and pathological temperature distributions, as all heating areas can be controlled individually to a desired target surface temperature.

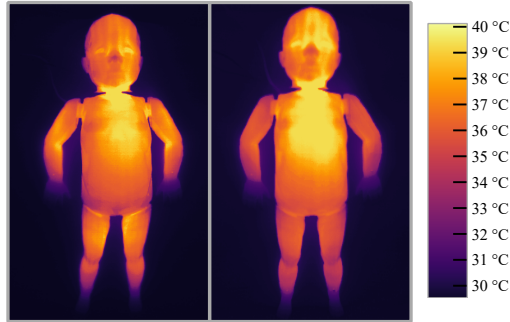


Abbildung 5.19: Thermal image of a physiological temperature distribution (left) and a central-peripheral temperature gradient for the extremities (right).

5.4 Summary

This chapter presented the development of a thermal phantom that accurately simulates the thermoregulation of a premature infant. Based on a 3D model of an infant, the phantom was remodeled to represent a premature infant (38 cm in length) at 29 weeks of gestation.

3D printed scaffolds were designed to house carbon fiber heating elements and Pt1000 temperature sensors used as heating components. This setup allows for independent temperature control and measurement in 13 different body zones. A silicone skin layer, modified with 16.7 % boron nitride for improved heat distribution, encased the phantom. 3D-printed ball-and-socket joints connected the limbs and head to the torso for realistic positioning and movement. The heating and sensing elements were controlled by an external circuit board connected to a computer running Matlab/Simulink. PI controllers regulated each heating zone, allowing for both constant and dynamic temperature profiles.

The phantom successfully maintained a uniform, physiological skin temperature with minimal deviation. Critically, it could also reproduce pathological

scenarios such as increased central-peripheral temperature differences, demonstrating its value for testing non-contact monitoring systems and control algorithms.

While the current phantom design provides a robust platform, future optimizations could enhance its functionality. The integrated circuit board heats up during operation, affecting the measurement and temperature behavior of the phantom. In future work, a circuit board with lower power dissipation could be developed, as well as the positioning of the circuit board in the phantom body could be adjusted. Incorporating a real-time, responsive model of infant thermoregulation would further enhance the physiological realism of the phantom. Finally, automating the movement of body parts through motorized joints would facilitate data collection for non-contact measurement systems.

These improvements, combined with the existing functionality of the phantom, will contribute to the development of robust and reliable non-contact skin temperature control systems for neonatal incubators.

6 Contactless Skin Servo Control for Neonatal Incubators

Maintaining stable skin temperature is critical for the health and survival of premature infants. Contactless skin temperature control is a promising approach to improve the safety and comfort of premature infants in neonatal incubators. Based on the previous chapters about temperature monitoring and a realistic physical phantom, this chapter presents the development and validation of a novel contactless skin servo control system for neonatal incubators. The system integrated a Hardware-in-the-Loop test bench equipped with automated infrared thermography for skin temperature measurement and a control algorithm that adjusts the incubator air temperature based on real-time skin temperature feedback. Two PI-based control strategies were designed and evaluated through both Software-in-the-Loop simulations and Hardware-in-the-Loop experiments, focusing on disturbance rejection (e.g., incubator door opening) and impaired thermoregulation scenarios (e.g., simulated hypothermia). The results demonstrate the effectiveness of both controllers in maintaining the target skin temperature while minimizing air temperature fluctuations. This study provides a proof-of-concept for contactless skin servo control, allowing for future clinical trials and the advancement of personalized thermoregulation in neonatal care. Parts of this chapter have already been published in peer-reviewed journals and conferences [VGW⁺24], [VLL24].

6.1 Motivation and Overview

Incubators provide a controlled microenvironment that allows precise adjustment of air temperature, relative humidity and oxygen levels. This microclimate is critical for premature infants with impaired thermoregulation and enables optimal growth and development. Early research by Silverman et al. [SFB58] and Jolly et al. [Jol61] demonstrated the significant impact of incubator care on survival rates of hypothermic infants.

Besides air temperature control, modern incubators offer advanced features such as skin servo control. While this advanced technology is cable-based, it allows direct control of the infant's skin temperature to a desired set point, further improving temperature management and overall comfort [LW16].

Monitoring skin temperature at a single site, typically the liver region, allows central skin temperature to be controlled. However, this approach may ignore peripheral skin temperature fluctuations. These fluctuations, which often precede central temperature changes due to factors such as peripheral vasoconstriction (induced by hypothermia), may expose the infant to thermal stress despite central temperature control. This limitation has been highlighted as a

criticism of local skin temperature control [DLRR⁺85, LP15].

Thomas et al. conducted two studies comparing air temperature control with skin temperature control in premature infants [TB99, Tho03]. In both studies, two groups of infants were monitored for 24 hours, one receiving skin temperature control and the other receiving air temperature control. The results demonstrated that skin temperature control provided more stable skin temperatures, while, at the same time, the air temperature fluctuated more.

In addition, a systematic analysis of two studies revealed that skin temperature control significantly increased the survival rate of low-birth-weight infants compared with care at a constant air temperature [Sin02]. This finding further highlights the potential benefits of skin temperature control for this vulnerable population.

About a decade ago, Deguines et al. [DDP⁺12] reported that skin temperature control was the preferred method in French hospitals, particularly in the days immediately following birth. Skin temperature control is now widely used in infant incubators. However, despite its widespread use, the precise implementation of commercially available systems remains largely unclear. Companies such as GE Healthcare are advertising their cascade control algorithm, which is designed to minimize temperature fluctuations in infants [Hea22]. In addition, Dräger Medical GmbH has patented a four-stage cascade control system specifically for skin servo control [Dra19]. Some researchers have explored the use of fuzzy control, which takes into account both current skin and air temperatures to minimize variations in both [RMH09]. These approaches demonstrate the ongoing quest to optimize skin temperature control for infant care.

Improving the comfort and development of premature infants who require skin temperature control in the incubator remains an active area of research. While skin temperature control is widely used, contact-based sensors pose a potential risk to the delicate skin of newborns. These sensors can cause damage and increase the risk of neonatal sepsis, a common complication in premature infants [SG13].

Non-contact temperature measurement in the incubator offers a promising solution to avoid potential skin damage associated with contact-based sensors. This approach meets the desires of parents and caregivers who see it as a way to improve parent-infant bonding and newborn comfort [BBCL17]. Infrared thermography is a key non-contact temperature measurement technique currently under research. Here, the goal is to further improve IRT to achieve levels of reliability and accuracy comparable to existing methods [KDHDS⁺17].

Continued advances in infrared thermography for neonatal incubators, as presented in Chapter 4, hold the potential for contactless skin servo control in the future, eliminating the need for attached sensors. This could significantly improve the comfort and development of premature infants [HHA⁺22b].

This chapter presents the novel approach of contactless skin servo control for neonatal incubators by applying both Software-in-the-Loop (SIL) and HIL setups, to address the issue of impaired thermoregulation in premature infants. This approach allows for the design, testing, and implementation of control strategies for contactless skin servo control prior to clinical application.

For this purpose, the neonatal phantom, as described in Chapter 5, was used, which allows the simulation of both physiological and pathological temperature distributions. In addition, the IRT-based temperature measurement system presented in Chapter 4 will be integrated to allow automated temperature extraction from the physical phantom. The resulting test bench can replicate various thermal scenarios, helping to explore the limits and potential of skin temperature control and perform extensive testing prior to clinical validation.

To demonstrate the functionality of contactless skin servo control, initial controllers have been designed and evaluated through simulation and real-world testing in the HIL test bench. This chapter provides a valuable resource for exploring the limits and possibilities of skin temperature control, paving the way for future clinical trials.

6.2 Hardware-in-the-Loop Setup

The following section describes the setup of the HIL test bench, which also enables open-loop measurements to be recorded. These measurements are important for system modeling and identification, enabling the development of accurate models for controller design and simulation. Figure 6.1 provides a schematic overview of the HIL test bench setup.

The experimental setup consisted of an incubator, the previously presented neonatal phantom, a multimodal camera system (including an RGB camera and an infrared thermal camera) with a reference body, and a laptop. The laptop served as the central control unit, connected to all components and handling data acquisition and processing.

The Thermocare Vita incubator (Weyer GmbH) was used. This double-walled convection incubator allows independent control of air and patient temperatures. Air temperature can be measured at two points within the chamber and the temperature displayed is the arithmetic mean. The incubator also has two connectable temperature sensors for measuring central and peripheral skin temperature [Wey23]. The incubator is equipped with a serial RS232 interface for temperature reading and air temperature target setting. The internal air temperature controller determines the temperature based on the set target. Notably, the standard incubator body is not transparent to infrared radiation. As described in Chapter 4, a ClearIR-4-P infrared window (Iriss) was integrated

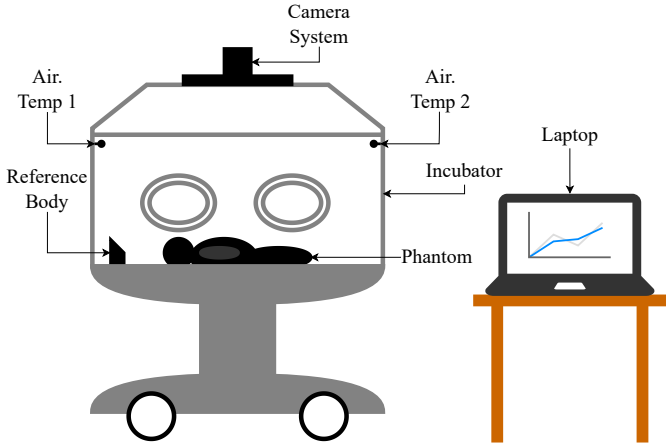


Abbildung 6.1: HIL test bench setup consisting of the incubator, the neonatal phantom, the camera system (including a reference body) and a laptop as the central control unit.

into the body of the incubator, which guarantees transparency for long-wave infrared rays ($8\text{--}14\text{ }\mu\text{m}$). The setup with neonatal phantom and reference body can be seen in Figure 6.2.

6.2.1 Automated Temperature Measurement

For non-contact temperature measurement, the spatial temperature monitoring framework described in Chapter 4 was implemented. The same PI 640 infrared camera (Optris) was used. However, a *IMX335* RGB camera (Waveshare) was integrated here for skin segmentation.

Reliable skin temperature measurements using IRT were obtained by applying the developed live correction algorithm. This algorithm is based on an integrated reference body that is passively heated to incubator temperature and positioned within the field of view of the infrared camera. The reference body allows compensation for the influence of incubator air temperature, humidity, and other factors on the accuracy of IR imaging. A more detailed theoretical explanation of the temperature correction algorithm can be found in Chapter 4.

After temperature correction, the deep learning model introduced in Chapters 3 and 4 was used for semantic segmentation. Here, the U-Net based neural



Abbildung 6.2: Test bench with neonatal phantom and reference body inside the incubator.

network used RGB image data to predict the skin of different body parts. The evaluation proved successful segmentation of all body parts.

The neural network for this chapter was specifically trained on manually labeled images of the neonatal phantom, ensuring high segmentation accuracy for the HIL setup. Each image pixel was automatically classified into one of five classes: head, torso, arms, legs, and background. This classification approach ensures reliable temperature measurement even when the infant is moving. The generated mask from the RGB camera was then mapped to the image space of the infrared camera using a homography transformation. This allowed the determination of averaged temperature values for the four body regions. Background pixels were ignored.

The described temperature correction and segmentation processes were implemented in Python and executed in real-time, providing the necessary data for the subsequent skin temperature control at a rate of one second. A detailed explanation of the neural network architecture, training, and evaluation can be found in Chapters 3 and 4.

6.2.2 Neonatal Phantom

The thermal phantom introduced in Chapter 5 was used as a realistic representation of a premature infant in an incubator. The 38 cm phantom closely approximated the body of a premature infant at approximately 29 weeks gestation. This phantom can adjust temperature values and trajectories, allowing the study of both physiological and pathological temperature profiles (see Fi-

gure 6.3).

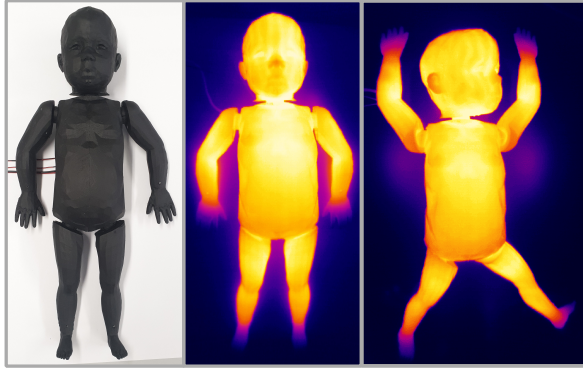


Abbildung 6.3: Neonatal phantom with different body postures (color and thermography).

As mentioned before, the thermal phantom was manufactured with a 3D-printed base containing an embedded circuit board for power electronics and serial communication via USB. To improve heat distribution and emissivity, the phantom was then covered with silicone skin and painted with black acrylic paint. The phantom body was segmented into 13 independently heated areas. Each area is equipped with a Pt1000 temperature sensor for accurate temperature measurement. A serial interface allows temperature data to be recorded and PWM values to be set for each area. A comprehensive description of the phantom's software, hardware, and underlying theoretical principles is presented in Chapter 5.

The temperature control of the thermal phantom was implemented in the *SIMULINK* graphical simulation environment (The Mathworks Inc.). PI controllers with saturation were used to control the temperature of the phantom. The manipulated variable, which corresponds to the PWM value set by the embedded circuit board, was limited to positive values and has individual upper limits for each segment. Importantly, this phantom controller differed from the skin servo controller, which will be described in the subsequent section.

In addition to the temperature set point, the phantom controller allowed the selection of a desired control mode. Three different control modes, detailed in Table 6.1, have been implemented and could be selected via the test bench.

Control mode **1** used a closed-loop configuration where the phantom controller dynamically sets the PWM value. Switching from mode **1** to mode **0**

Tabelle 6.1: Overview of the different phantom control modes.

Control Mode	Explanation
0	Manipulated variable (PWM) is maintained at a constant value (averaged over the last five minutes)
1	Phantom controller active
2	Phantom heating off

held the PWM value constant at the average value over the last 5 minutes.

Notably, the PWM value may not be the same when switching from mode 1 to mode 0 despite a constant temperature set point (e.g., 37°C). This is because the phantom heating process is affected by the incubator temperature. Suppose the incubator temperature is 31°C, which is well below the phantom target temperature of 37°C. In this case, the phantom will require more internal heating to reach the desired temperature, resulting in a higher PWM value. In contrast, if the incubator temperature is 34°C, which is closer to the phantom's target temperature, the phantom will require less internal heating. This would result in a lower PWM value. This dynamic adjustment of the PWM value, influenced by the incubator temperature, allows the phantom to simulate premature infants with different metabolic rates.

Control Mode **2** corresponds to a deactivated heating state, with the PWM value set to 0. When switching to Control Mode **1**, the integrator state was reset to 0, ensuring a consistent starting point for the control loop.

In addition, the phantom is capable of simulating the impaired thermoregulation of premature infants. This was accomplished by an adjustable heat factor f_{var} , which modified the constant PWM value in control mode 0. By manipulating this heat factor, either hyperthermia or hypothermia could be simulated. A higher heat factor increased the heat produced by the phantom, simulating hyperthermia, while a lower heat factor reduced the heat production, mimicking hypothermia. This capability provided a method for investigating the mechanisms of impaired thermoregulation in premature infants and exploring potential solutions to these problems.

It is important to note that the internal circuit board of the phantom generated additional heat due to the use of power transistors. This heat source reaches a steady state after a certain period of time, effectively acting as a constant thermal source. As a result, reducing the power supplied to the heating elements, for instance, by 50%, leads to a less significant reduction in the phantom's total heat output.

6.2.3 Software of the Test Bench

In addition to controlling the neonatal phantom, the Simulink software environment was also used to manage the HIL test bench. This centralized control allowed a single HIL computer to control the entire test setup, including skin servo control, the neonatal phantom, and automated temperature measurement using an IRT camera. A schematic representation of the key components of the test rig is given in Figure 6.4.

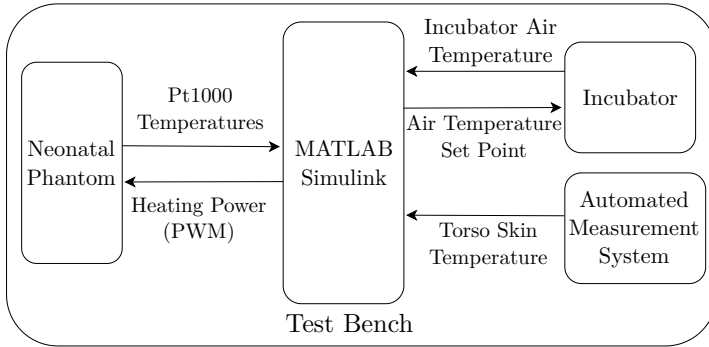


Abbildung 6.4: Functional overview of the HIL test bench.

The torso temperature measured obtained by the contactless temperature measurement system was defined as the skin temperature in this chapter. A new set point for the incubator temperature $T_{\text{Inc,set}}$ was then calculated based on the resulting control error and sent to the incubator.

6.3 Control System

The design of a skin servo controller $G_{C,\text{Skin}}$ required the modeling and identification of the corresponding control system. Figure 6.5 provides the control system structure adopted in this study.

Following the established contact-based skin servo control systems [Hea22, Dra19], a cascaded control strategy has been implemented. This design features an inner loop to control the incubator air temperature consisting of a controller $G_{C,\text{Air}}$ and a plant $G_{P,\text{Air}}$. Potential disturbances to this inner loop can be caused by opening the incubator's hand ports.

In addition, the plant $G_{P,\text{Skin},\nearrow}$ represents the time-variant heat dynamics of the infant, composed of the convective heat transfer $\dot{Q}_{\text{Convection}}$ between the

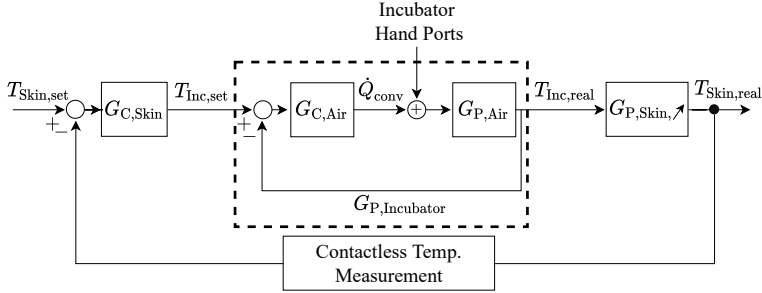


Abbildung 6.5: Cascaded control system with an inner loop for incubator air temperature control and an outer loop for skin temperature control with a time-varying plant.

incubator air and the infant, and the time-variant heat production $\dot{Q}_{heat,var}$ of the infant. As the inner loop elements ($G_{C,Air}$ and $G_{P,Air}$) are proprietary components of the commercial incubator design and not available in detail, the inner loop is represented by a combined system $G_{P,Incubator}$.

6.3.1 System Modeling

To simulate the previously proposed control system, the $G_{P,Incubator}$ plant and the dynamics of the infant were modeled and identified. Due to the active heating and passive cooling capabilities of the incubator, the dynamics of the controlled system $G_{P,Incubator}$ exhibited significant variations depending on whether the incubator temperature was rising or falling. Therefore, the system dynamics $G_{P,Incubator}$ were split into two separate models: $G_{P,Incubator,up}$ for increasing incubator temperature and $G_{P,Incubator,down}$ for decreasing incubator temperature.

The plant $G_{P,Incubator}$ was modeled using Matlab/Simulink. A switching logic was implemented to dynamically select between the two plant models $G_{P,Incubator,up}$ and $G_{P,Incubator,down}$ depending on the difference between the set point temperature $T_{Inc,set}$ and the actual temperature $T_{Inc,real}$. If the set temperature exceeds the actual temperature, $G_{P,Incubator,up}$ is activated. Likewise, $G_{P,Incubator,down}$ is activated when the current temperature exceeds the set point. To avoid frequent switching at small temperature differences, a threshold-based switching logic, similar to a Schmitt trigger, was implemented.

A variable state space model was implemented in Simulink to ensure a continuous transition between the two dynamic models. State-space representations were developed for both systems. Upon activation of the switching logic, the

corresponding state space matrices are swapped to reflect the required dynamics, ensuring consistent states and a continuous transition. In addition, an initial value for $T_{\text{Inc,real}}$ can be specified at the start of the simulation.

A thermal model $G_{\text{P,Skin},\nearrow}$ representing the dynamics of the infant was developed using Matlab Simscape. The heat balance of the modeled infant over a time interval dt can be expressed by the following equation:

$$\dot{Q}_{\text{heat,var}} - \dot{Q}_{\text{Convection}} = m_{\text{infant}} \cdot c_{\text{infant}} \cdot \frac{dT_{\text{Skin,real}}}{dt}, \quad (6.1)$$

where c_{infant} and m_{infant} are the specific heat capacity and the weight of the infant, respectively.

The heat exchange between the infant and the incubator air space is influenced by the incubator air temperature, T_{Inc} . For this model, a purely convective heat transfer was assumed. Thus, the heat exchange between air and skin was simplified to a linear dependence on the temperature difference. The convective heat transfer can be estimated by:

$$\dot{Q}_{\text{Convection}} = A_{\text{Skin,conv.}} \cdot h_c \cdot (T_{\text{Inc,real}} - T_{\text{Skin,real}}), \quad (6.2)$$

where h_c represents the convective heat transfer coefficient for this process.

A premature infant with a weight of $m_{\text{infant}} = 1.5 \text{ kg}$ was assumed in the model. Based on [LW16], the effective skin area for convective heat transfer can be calculated to:

$$A_{\text{Skin,conv.}} = 0.1 \cdot (m_{\text{infant}})^{\frac{2}{3}}. \quad (6.3)$$

Instead of relying on literature values, the convective heat transfer coefficient h_c , the specific heat capacity c_{infant} , and the heat production $\dot{Q}_{\text{heat,var}}$ were determined through a parameter identification process. This approach ensured good consistency between the simulation results and the phantom measurements obtained on the test bench.

The phantom used in the HIL test bench had a constant heat production generated by its internal circuit board and a variable heat production controlled by a laptop. In order to reproduce this behavior in the thermal model, the heat production was also divided into a constant and a variable component:

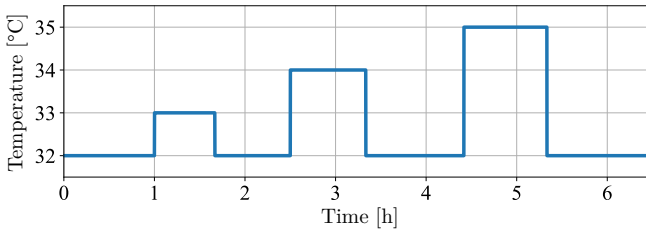
$$\dot{Q}_{\text{heat,var}} = \dot{Q}_{\text{heat}} \cdot f_{\text{const}} + \dot{Q}_{\text{heat}} \cdot (1 - f_{\text{const}}) \cdot f_{\text{var}} \quad (6.4)$$

The variable component of heat production, $\dot{Q}_{\text{heat}} \cdot (1 - f_{\text{const}})$, can be modified by the heat factor f_{var} . This allows the simulation of thermoregulatory pathologies (hypothermia or hyperthermia) similar to the neonatal phantom.

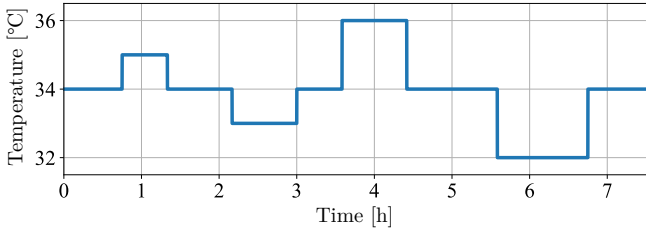
Finally, to replicate the measurement noise of the test bench temperature sensor, band-limited white noise with a power spectral density of $0.01 \text{ }^\circ\text{C/Hz}$ was added to the simulated skin temperature $T_{\text{Skin,real}}$.

6.3.2 System Identification

The system identification was performed in two steps. First, the incubator plant models, $G_{P,incubator,top}$ and $G_{P,incubator,bottom}$ were estimated. Then the parameters of the thermal model were identified. To obtain sufficient data for system identification, several step responses were recorded by adjusting the set point of the incubator air temperature $T_{Inc,set}$. These included step responses with amplitudes ranging from 1 °C to 3 °C and a hopping sequence around a set point in the range ± 2 °C. The corresponding set point sequences for the incubator air temperature $T_{Inc,set}$ are illustrated in Figure 6.6.



(a) Chosen temperature set point profile between 32 °C and 35 °C



(b) Chosen hopping temperature profile around a set point of 34 °C

Abbildung 6.6: Set point sequences (a) and (b) of incubator air temperature set point $T_{Inc,set}$ for system identification.

Several step responses and responses to the hopping sequences were recorded with initial temperatures of 32 °C and 34 °C (step response) and 34 °C and 36 °C (hopping sequence). Simultaneously, the neonatal phantom was placed in the incubator and heated to 37 °C using its internal phantom control.

To enable the design of a robust controller, both incubator plant models, $G_{P,Incubator,up}$ and $G_{P,Incubator,down}$, were assumed to follow first-order dyna-

mics with uncertain gain and time constant (cf. equations (6.5)-(6.6)).

$$G_{P,Incubator,up}(s) = \frac{K_{Incubator,up}}{1 + T_{Incubator,up} \cdot s} \quad (6.5)$$

$$G_{P,Incubator,down}(s) = \frac{K_{Incubator,down}}{1 + T_{Incubator,down} \cdot s} \quad (6.6)$$

The recorded data was further segmented into individual step responses as depicted in Figure 6.7.

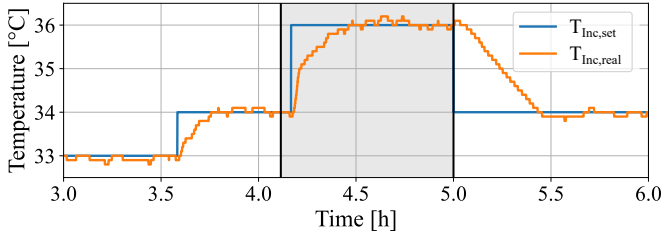


Abbildung 6.7: Segmentation of recorded step response data into individual step responses.

This generated a total of 40 individual step responses for $G_{P,Incubator,up}$ and 37 for $G_{P,Incubator,down}$, which were then used to estimate the four uncertain parameters: $K_{Incubator,up/down}$ and $T_{Incubator,up/down}$. Outliers due to high ambient temperatures during the summer months and interrupted measurements were excluded from the analysis.

Parameter estimation was performed using a nonlinear least squares method, minimizing the following objective function:

$$\min V = \sum_{t=t_1}^{t_2} (T_{Inc,real}(t) - T_{Inc,sim}(t))^2 \quad (6.7)$$

$$\text{where } V = f(K_{Incubator,up/down}, T_{Incubator,up/down}),$$

with $T_{Inc,real}$ describing the measured air temperature of the incubator, $T_{Inc,sim}$ the simulated output of the first-order dynamics (cf. equations (6.5) and (6.6)), and f the model implemented in Matlab Simulink.

Table 6.2 provides an overview of the identified uncertain parameters, including their mean and standard deviation.

Next, the parameters of the thermal model were identified. The previously recorded step responses, obtained with the neonatal phantom placed in the

Tabelle 6.2: Uncertain parameters of $G_{P,Incubator}$.

Parameter	Mean Value	STD	Value Range
$K_{Incubator,up}$	1.003	0.003	[1, 1.012]
$K_{Incubator,down}$	0.995	0.005	[0.984, 1.003]
$T_{Incubator,up}$	316 s	100 s	[105 s, 475 s]
$T_{Incubator,down}$	866 s	370 s	[294 s, 1681 s]

incubator and heated to 37°C, were used for this purpose. The parameters were estimated by minimizing the following objective function:

$$\min S = \sum_{t=t_1}^{t_2} (T_{Skin,real}(t) - T_{Skin,sim}(t))^2 \quad (6.8)$$

$$\text{where } S = f(h_c, c_{infant}),$$

with $T_{Skin,real}$ describing the measured skin temperature of the infant, $T_{Skin,sim}$ the simulated output of the infant model, and f the model implemented in Matlab Simulink.

Using the step response data, the convective heat transfer coefficient (h_c) and the specific heat capacity of the infant (c_{infant}) were estimated to be $h_c = 10 \frac{W}{m^2 \cdot K}$ and $c_{infant} = 1400 \frac{J}{K \cdot kg}$. A direct comparison of the estimated heat transfer coefficient with typical values is not viable due to the assumptions of the model. In addition, the estimated specific heat capacity is about 47% of the physiological value reported in [FBNK98]. This deviation is caused by the approximations of the model and the specific differences between the phantom and a living infant.

Finally, the parameters related to heat production $\dot{Q}_{heat,var}$ were identified to characterize the different heat sources within the thermal phantom. Measurements were made with the phantom heated to 37°C at a fixed incubator temperature of 32°C. Afterwards, the phantom heat factor f_{var} was varied between 0.5 and 1. This variation allowed the separation of the different heat sources within the phantom. The heat production parameters were then optimized by minimizing the following objective function:

$$\min M = \sum_{t=t_1}^{t_2} (T_{Skin,real}(t) - T_{Skin,sim}(t))^2 \quad (6.9)$$

$$\text{where } M = f(\dot{Q}_{heat}, f_{const}),$$

with $T_{\text{Skin,real}}$ describing the measured skin temperature of the infant, $T_{\text{Skin,sim}}$ the simulated output of the infant model, and f the model implemented in Matlab Simulink.

The identification process provided estimations of $\dot{Q}_{\text{heat}} = 6.6 \text{ W}$ and $f_{\text{const}} = 0.48$ for heat production and constant heat factor, respectively. This allows for the simulation of the same scenarios of impaired thermoregulation (hypothermia or hyperthermia) both in the model and on the test bench using the phantom.

6.3.3 Control Design

Two initial skin servo controllers, $G_{C,\text{Skin}}$, were designed based on the identified plant models. Disturbance suppression was prioritized in the design process, aiming for a phase margin of $\alpha_{rd,\text{set}} = 60^\circ$ to ensure robust stability and adequate performance. A simple PI controller was implemented, defined as

$$G_{C,\text{Skin}}(t) = K_p \cdot e(t) + K_i \cdot \int_0^t e(\tau) d\tau, \quad (6.10)$$

$$\text{with } e(t) = T_{\text{Skin,set}} - T_{\text{Skin,real}}.$$

The PI controller parameters were initially tuned using Matlab to achieve the desired phase margin of 60° . Further manual adjustments were made to optimize the control performance, resulting in $K_p = 0.69$ and $K_i = 6.6 \cdot 10^{-4}$.

As an alternative approach, a gain-scheduled PI controller was implemented. In this design, K_i was fixed at 10^{-3} , while K_p was a function of the absolute value of the control error (see Figure 6.8).

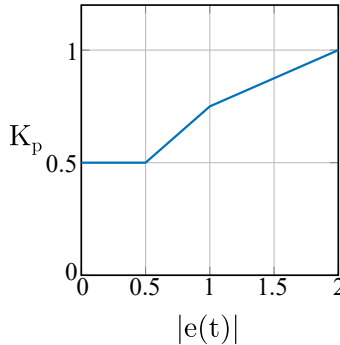


Abbildung 6.8: Gain scheduling of K_p as a function of the control error $|e(t)|$.

This approach was chosen to improve the controller's response to larger control errors, allowing it to reach the set point more quickly. At the same time,

reduced gain for smaller errors minimizes the effect of measurement noise disturbances.

Both controllers were discretized using a forward difference quotient. A sampling time of $\Delta T = 60 \text{ s}$ was chosen for the controllers. Afterwards, the open-loop system G_0 was analyzed using a Bode plot, with G_0 defined as:

$$G_0 = G_{C,\text{Skin}} \cdot G_{P,\text{Incubator}} \cdot G_{P,\text{Infant},\text{lin.}} \quad (6.11)$$

where $G_{P,\text{Infant},\text{lin.}}$ describes the linearized plant of the infant model (cf. equation (6.1)).

Figure 6.9 provides the Bode plot of G_0 with the PI controller for different values of the uncertain parameter $T_{\text{Incubator}}$ in the range $[105 \text{ s}, 1681 \text{ s}]$.

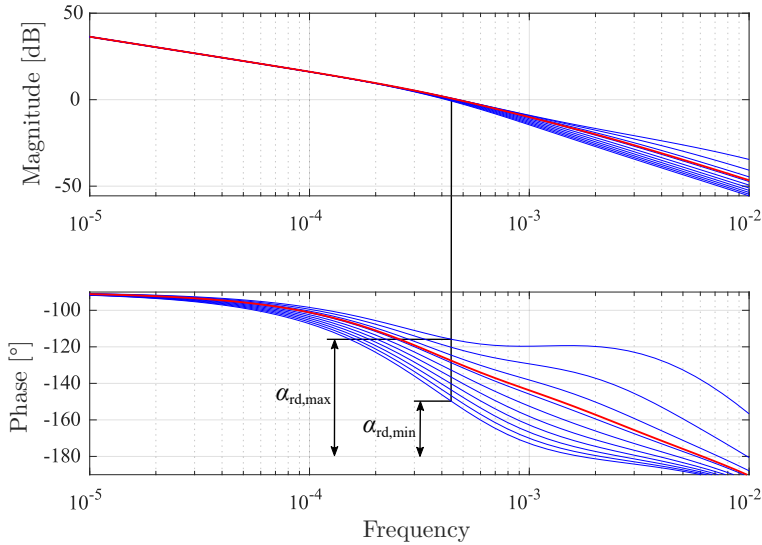


Abbildung 6.9: Bode plot of the open-loop system with a simple PI controller, visualizing the influence of variations due to the uncertain parameter $T_{\text{Incubator}}$. The nominal system is given in red.

The Bode plot clearly demonstrates the impact of the uncertainties in the controlled system, $G_{P,\text{Incubator}}$. The phase margin of the controlled system varies between 33.3° and 64.8° due to these uncertainties. Despite this variation, the control loop remains stable for the modeled parameter uncertainty.

In addition, a Bode plot analysis of the open loop system with the gain-scheduled PI controller was performed considering the minimum and maximum

values of K_P (0.5 and 1 as given in Figure 6.8). The corresponding Bode plots for both values of K_P are provided in Section A.2 of the Appendix for reference. This analysis revealed a minimum phase margin of $\alpha_{rd,min} = 13.4^\circ$, indicating that the control system is operating close to the stability limit, potentially leading to more aggressive control behavior.

Initial tests indicated that frequent changes in the incubator set temperature $T_{Inc,set}$ resulted in increased air temperature oscillations. To address this issue, an additional 0.2°C threshold was implemented for the air temperature set point. This threshold allows a set point change only if the new set point is at least 0.2°C different from the previous one.

6.4 Results

The concept of contactless skin servo control was evaluated in two steps. First, a classic disturbance rejection scenario was examined by evaluating the effects of opening and closing incubator hand ports on both skin temperature and incubator air temperature. Second, the designed controllers were evaluated under simulated thermoregulatory challenges commonly encountered by premature infants. These challenges were implemented by adjusting the previously defined heat factor f_{var} . This way, hypothermia/hyperthermia was simulated. Details of the experimental setup can be found in Section 6.2.

The modeled control system was first evaluated using Software-in-the-Loop simulations before being transferred to the test bench for subsequent HIL testing.

6.4.1 Disturbance Rejection

For the evaluation of the classical disturbance rejection of the skin control system (PI controller), two of the four hand ports of the incubator were opened. This 6.3-hour study began with a one-hour heating of the phantom (control mode 1) at a constant incubator temperature of 32°C . The control mode of the phantom was then set to 0, while the heating factor was kept at 1 for the rest of the experiment. At the same time, the skin servo controller was activated. The hand ports of the incubator were then opened for 15 minutes and then again for 30 minutes.

Figure 6.10 presents the resulting skin temperature, as well as the actual and target incubator air temperatures. The two phases of open incubation hand ports are highlighted. The first hour of the simulation, characterized by constant incubator temperature and steady-state skin temperature, has been excluded for clarity.

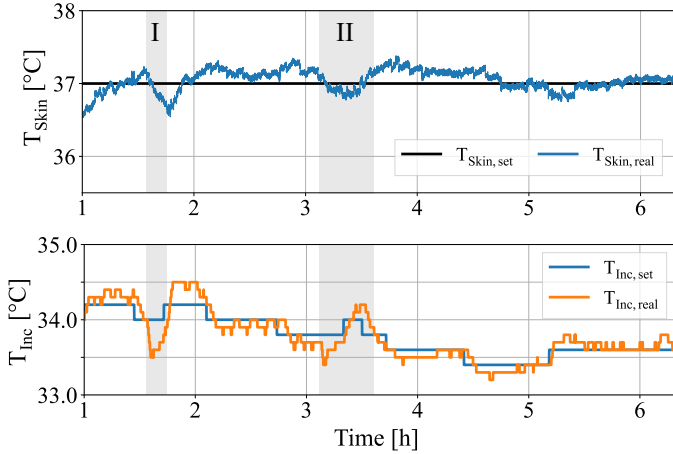


Abbildung 6.10: Disturbance rejection performance of the cascaded skin controller during two instances of incubator hand port opening (I and II).

A decrease in both skin and air temperature was observed during both phases of opening the incubator hand port. While the temperatures decreased by 0.5°C during Phase I, only a small temperature change was observed during Phase II. In particular, the set point air temperature, representing the control variable of the skin servo controller, experienced a change of only 0.2°C in both phases. In contrast, the actual air temperature exhibited a significant overshoot after the initial drop when the incubator hand ports were closed again.

6.4.2 Impaired Thermoregulation

The designed controllers were evaluated for their ability to meet the thermoregulatory needs of premature infants, with a particular focus on simulated scenarios of impaired thermoregulation, including both hypothermia and hyperthermia. To implement this time-varying scenario, the heat factor variation presented in Figure 6.11 was applied.

This seven-hour long scenario started with an one-hour heating of the phantom (control mode 1) at a constant incubator air temperature of 32°C . The phantom control was then deactivated (control mode 0) and the heating factor was kept constant for 1.5 hours. At the same time, the skin servo controller was activated to control the skin temperature to the set point of 37°C . After 2.5 hours, the heat factor was decreased to 0.5 over a period of 1 hour and held at

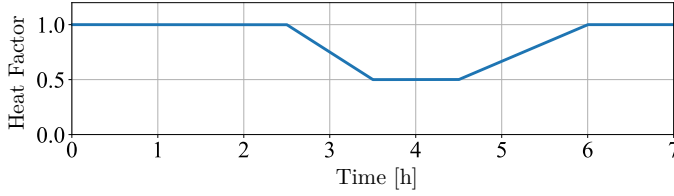


Abbildung 6.11: Heat factor f_{var} scenario for control system evaluation.

that value for another hour, simulating impaired thermoregulation in the form of hypothermia. The heat factor was then gradually increased back to 1 over 1.5 hours and finally held constant at 1 for the remaining hour, simulating a return to a healthy physiological state.

Two quality criteria have been defined to evaluate the performance of the skin servo controllers. The first criterion, referred to as the set point range (SPR) criterion, measures the percentage of total measurement time during which the skin temperature remains within a range of $\pm 0.25^\circ\text{C}$ around the set point of 37°C . Specifically, the SPR criterion is considered true when

$$SPR : |T_{\text{Skin,set}} - T_{\text{Skin,real}}(t)| < 0.25^\circ\text{C} \quad (6.12)$$

The second criterion is the mean square error (MSE) between the set point and the actual skin temperature:

$$\text{MSE} = \frac{1}{N} \sum_{t=t_1}^{t_2} (T_{\text{Skin,set}} - T_{\text{Skin,real}}(t))^2, \quad (6.13)$$

with N being the number of temperature values during times t_1 and t_2 .

Software-in-the-Loop Experiment

The SIL experiment allows a preliminary examination of the controller's performance under simulated conditions of impaired thermoregulation, providing a basis for subsequent HIL testing. The simulation, which uses a variable heat factor f_{var} , evaluates the controller's response to various levels of simulated impaired thermoregulation, including induced hypothermia and hyperthermia. Figure 6.12 illustrates the results of this experiment.

The simulated skin temperature and the actual incubator air temperature are provided, comparing the dynamics of two closed-loop temperature control systems and an open-loop system without skin temperature control. Additionally, the desired set temperature of $T_{\text{Skin,set}} = 37^\circ\text{C}$ is included for reference. The heat factor is also given for better visualization.

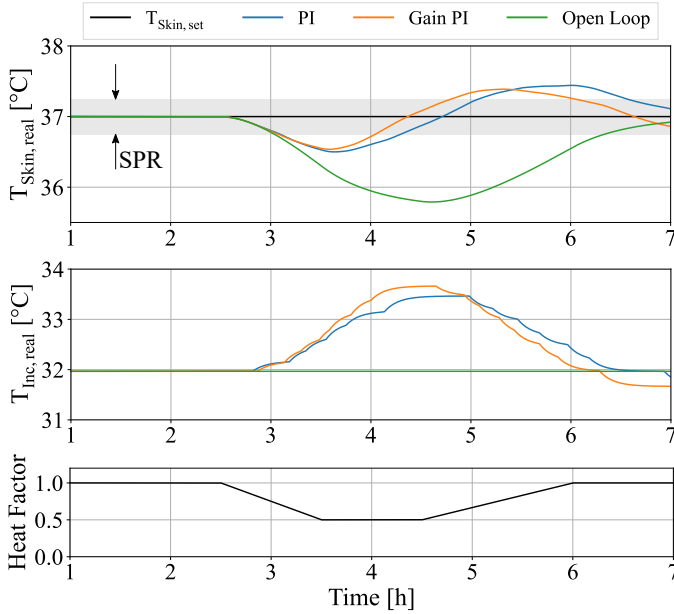


Abbildung 6.12: Comparison of simulated controller responses to a simulated scenario of impaired thermoregulation. The grey band represents the 0.25 °C SPR.

The simulation started with initial temperatures of $T_{\text{Skin,real}} = 37^\circ\text{C}$ and $T_{\text{Inc,real}} = 32^\circ\text{C}$. The first hour of the simulation, characterized by constant incubator temperature and steady skin temperature, has been excluded for clarity. After one hour, a steady state skin temperature of approximately 37°C was reached. At this point, the skin controller was activated and successfully maintained the skin temperature at the 37°C setpoint. Notably, the skin temperature also remained stable at 37°C in the open-loop simulation prior to any change in the heat factor.

In the next phase, hypothermia was simulated by reducing the heat factor to 0.5, which resulted in a rapid decrease in skin temperature in all experiments. However, the enabled skin controllers effectively compensated for this temperature drop by increasing the air temperature. The gain-scheduled PI controller demonstrated a faster response than the simple PI controller in this scenario. Conversely, the skin temperature experienced a more significant decrease in the open loop experiment, reaching a minimum of 35.8°C . Increasing

the heat factor from 0.5 to 1 simulated a return to a physiological state for premature infants, causing both skin controllers to overshoot as they adapted to the change. Again, the gain-scheduled PI controller responded faster and reached the setpoint first.

Adding to the qualitative results, a quantitative evaluation of the simulation was performed using the previously defined criteria. The quantitative results are summarized in Table 6.3.

Tabelle 6.3: Quantitative evaluation of the simulation using MSE and SPR criteria, as defined in equations (6.12) and (6.13).

Controller	MSE	SPR
Open loop	0.438	44.99 %
PI control	0.076	55.30 %
Gain scheduled PI control	0.055	63.31 %

Both closed-loop systems demonstrated a significant advantage over the open-loop case. Consistent with the qualitative results, the gain-scheduled PI controller outperformed the simple PI controller in both criteria, exhibiting an 8 % improvement in the percentage of time within the defined setpoint range.

Hardware-in-the-Loop Experiment

After initial evaluation through Software-in-the-Loop simulation, the controllers were tested on the dedicated HIL test bench. This testing evaluated the ability of each controller to compensate for simulated thermoregulatory challenges, including hypothermia, using the novel contactless skin servo system. The results, which compare the different control strategies on both skin temperature (measured by an infrared thermal camera) and incubator air temperature, are provided in Figure 6.13.

In the end of the phantom heating process (after 1 hour), the skin temperature was observed to be slightly below 37 °C for all measurements. In the open-loop measurement, the temperature continued to rise despite a constant heat factor. In contrast, both the PI and gain-scheduled PI controllers effectively maintained a constant skin temperature of 37 °C.

During the simulated hypothermia phase, both skin controllers quickly compensated for the disturbance and prevented the skin temperature from dropping below 36.5 °C. The gain-scheduled PI controller again demonstrated a faster response in reaching the 37 °C set point. At the same time, the air temperature in the incubator increased from 32 °C to nearly 34 °C. In contrast, the skin temperature in the open-loop measurement dropped below 36 °C and did not

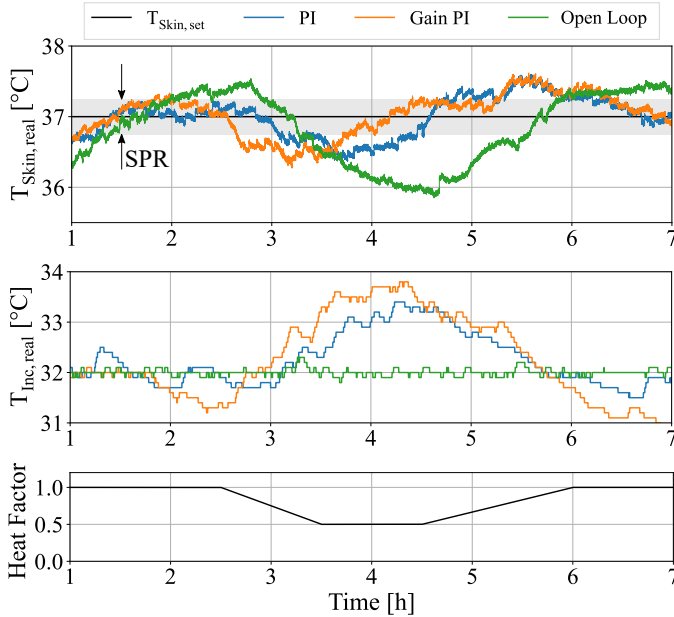


Abbildung 6.13: Controller performances for simulated impaired thermoregulation on the HIL test bench. The grey band represents the 0.25 C° SPR.

return to the set point until 5 hours later. During the simulated return to a physiological state (increasing the heat factor from 0.5 to 1), both control systems exhibited a similar brief overshoot before quickly returning to the 37 °C set point. Similar to the simulation results, a quantitative evaluation of the HIL experiment was performed and the results are summarized in Table 6.4.

Tabelle 6.4: Qualitative evaluation of the HIL experiment using MSE and SPR criteria.

Controller	MSE	SPR
Open loop	0.274	22.80 %
PI control	0.079	51.15 %
Gain scheduled PI control	0.088	59.22 %

In this specific HIL experiment, both controllers revealed a significant advantage over the uncontrolled skin temperature, achieving similar performance

levels. In contrast to the simulation results, the PI controller had a slightly lower MSE. However, the gain-scheduled PI controller outperformed it in terms of the percentage of time within the defined setpoint range, achieving an 8 % improvement.

At the end, a comparative evaluation was made between the measurements obtained on the test bench and the simulated results. Figure 6.14 presents this comparison, providing the response of both the gain-scheduled PI controller and open-loop case without skin temperature control.

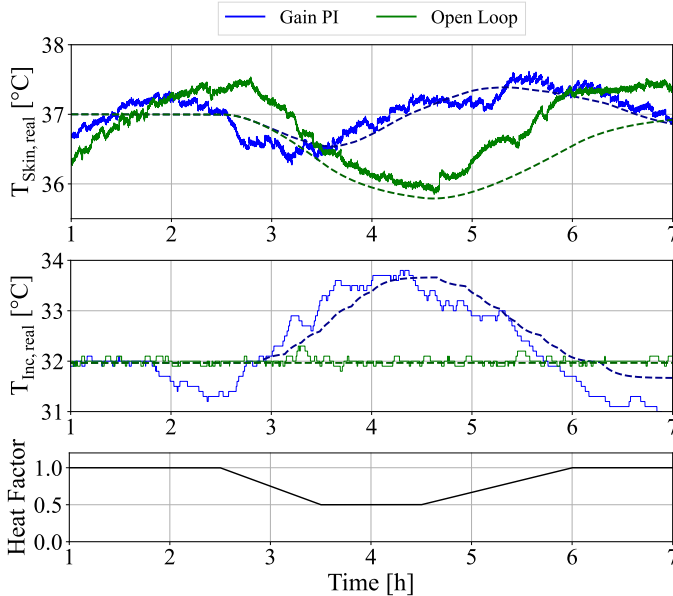


Abbildung 6.14: Comparison of simulated (dashed lines) and HIL test bench (solid lines) results for both the gain-scheduled PI controller and the open-loop system.

The most significant deviation between the simulation and HIL measurement occurred in the skin temperature after one hour. At this point, the simulation reached a steady state skin temperature of 37 °C with a fixed heat factor and a constant incubator temperature of 32 °C. However, the skin temperature measured on the test bench was slightly below 37 °C. Despite this difference, the temperature curves from both the simulation and the actual measurement demonstrated a similar trajectory during variations in heat production. This

alignment between simulation and HIL test bench results was observed for both the closed-loop and open-loop cases. The only notable deviation was a slightly higher skin temperature during the open-loop measurement on the test bench compared to the simulation.

6.5 Discussion

This chapter has successfully demonstrated the potential of contactless skin servo control for neonatal incubators using basic PI controllers. Both the simulated and experimental trials on the test bench achieved promising results, demonstrating the effectiveness of the skin controllers in minimizing disturbances, including the opening of incubator hand ports and scenarios of impaired thermoregulation.

The opening of incubator hand ports was effectively compensated by an increased air temperature. Specifically, the air temperature set point (the control variable of the skin controller) increased only slightly, while the actual air temperature experienced a significant overshoot after an initial decrease. This finding suggests that the disturbance can effectively be compensated by an inner air temperature loop within a cascaded control system, demonstrating the robustness of the system.

The direct comparison between simulation and HIL measurement indicates that the model developed in this study accurately replicates the test bench, including the phantom, and thus enables the generation of realistic simulation results. While the gain-scheduled PI controller demonstrated superior performance for both criteria in the simulation, the HIL measurements revealed a higher percentage of time within the setpoint range for the gain-scheduled PI controller. However, the simple PI controller had a slightly lower mean square error. The limited number of measurements and the lack of various pathological temperature scenarios require a cautious interpretation of these results. Nevertheless, this work serves as an important proof-of-concept for future investigations of non-contact skin servo control for neonatal incubators, advancing the state of the art towards more precise and personalized thermoregulation for premature infants.

Future research is needed to improve the robustness of the controller validation through additional experiments covering a wider range of temperature and heat production scenarios. The presented HIL setup enables this by providing flexibility in configuring various settings, which includes changes in phantom properties (simulating impaired thermoregulation and pathological scenarios), modifications in the applied temperature measurement technique, and adjustments in the implemented skin control algorithm. This functionality allows

comprehensive testing of the robustness, stability and dynamics of control algorithms, serving as a critical step in preparing and conducting clinical trials and obtaining regulatory approval. In addition, the setup can help to improve the regulatory review process itself.

However, certain limitations within the test bench require further investigation. Heat dissipation from the PCB within the phantom limits the ability to accurately reproduce the full range of thermal characteristics of premature infants. As a result, validation of control algorithms for very young premature infants remains a challenge. Future revisions of the circuit board components are planned to address the heat dissipation issue.

Furthermore, the phantom currently only includes heat loss due to convection, conduction, and radiation. This limitation hinders the accurate representation of the significant heat loss due to evaporation observed in real premature infants.

The non-contact automated temperature measurement system requires further validation. As clinical practice currently depends on single-point measurements in the liver region using skin-attached sensors, future research should investigate the most suitable metric for skin servo control when using spatially resolved temperature measurement based on IRT. In addition, the thermal phantom used in this study lacks the ability to move body parts, which differs from real-world scenarios. Movement of premature infants may introduce imaging artifacts and variations within the measurable skin area within the field of view. Therefore, further research is needed to determine the stability of specific temperature measurement metrics in relation to movement.

While this chapter used automatically extracted torso temperature, the setup allows for the inclusion of temperatures from the head, arms, or legs. The development of a central-peripheral temperature difference calculation that incorporates the centralizing effect observed during infection/sepsis could improve the control algorithm. This would allow for more personalized and precise control strategies tailored to specific physiological conditions. In addition, the current setup assumed a target torso temperature of 37 °C and focused primarily on establishing proof-of-concept for the HIL setup and contactless skin servo control. Future research is needed to determine the optimal skin temperature set point for different scenarios. This objective requires the development and validation of a model capable of estimating core body temperature from contactless skin temperature measurements. This model could be developed and validated through an observational clinical study, which can simultaneously assist in the evaluation of the automated temperature measurement system.

Future efforts in robust controller design should focus on developing a more sophisticated model of the control system. The current model relies on first-order lag elements and fundamental principles. A thermal compartment model

capable of simulating the complex dynamics of the incubator and premature infants could be advantageous. This model would allow the simulation of different thermal masses (air space, mattress, premature infant, incubator walls) and their heat transfer, thus providing a more realistic representation. In addition, modeling of the simulated premature infant with uncertain parameters such as skin area and metabolic heat rate could be incorporated to simulate different stages of premature birth, further enhancing the realism and predictive accuracy of the model.

6.6 Summary

The concept of contactless skin servo control for neonatal incubators was introduced in this chapter, highlighting the potential of this technology to improve thermoregulation and enhance patient safety. It presented the development of a HIL test bench that incorporates a physical neonatal phantom to simulate physiological and pathological thermoregulatory scenarios. This novel setup, combined with the development and identification of a comprehensive control system model, enables the testing of skin servo controllers through both simulation and real-world experiments.

While the presented system provides valuable insights, further optimization has the potential to advance the development of skin temperature controllers with non-contact temperature measurement.

The thermal phantom could be improved by reducing the heat dissipation from the PCB, especially at higher incubator temperatures. In addition, the development of a dynamic model of infant thermoregulation that responds to changes in skin and ambient temperature would allow for more realistic simulations. This improved model, coupled with a detailed thermal model of the incubator and the newborn, would allow for more thorough testing of the designed controllers. With these future optimizations, the existing HIL test bench will provide a robust platform for the design and validation of skin temperature controllers, paving the way for subsequent clinical trials. Further development could include more sophisticated controllers, such as adaptive or model-predictive controllers, that can respond to changes in the controlled system.

This research demonstrates the possibility and potential of contactless skin servo control for neonatal incubators, opening new avenues for future research and clinical applications. By addressing the identified limitations and pursuing the proposed future directions, this technology offers the potential to significantly improve neonatal care and the outcomes for premature infants.

7 Conclusion

Preterm birth, a major global health problem affecting millions of infants each year, poses a number of challenges to infant health and well-being. These vulnerable newborns, lacking the more robust thermoregulatory mechanisms of their full-term counterparts, are highly susceptible to temperature fluctuations. Such fluctuations can lead to a cascade of negative outcomes, including increased oxygen requirements, impaired lung function, hypoglycaemia and even death. Traditional methods of monitoring skin temperature, which rely on wired thermistors attached to the infant's skin, pose additional risks, including skin damage, irritation and increased risk of infection, especially in these fragile newborns. In addition, the presence of wires can interfere with parent-infant bonding, complicate medical procedures and restrict the infant's movement in the incubator, which can affect their physical development and emotional well-being.

This thesis aimed to address these limitations by exploring the potential of non-contact infrared thermography as a more comfortable, safe and informative alternative. By integrating deep learning and control algorithms, the research aimed to develop and validate a novel multimodal approach to contactless skin temperature management, ultimately contributing to improving the care and well-being of premature infants.

Chapter 3 aimed to develop and evaluate deep learning models for automatic infant body part segmentation, enabling accurate extraction of spatial temperature data from IRT images. A robust dataset of 600 RGB and 600 LWIR images of premature infants, carefully selected to cover different postures, movements and clinical scenarios, was meticulously annotated. The study compared different U-net-based network architectures known for their effectiveness in medical image segmentation tasks. The research explored the effectiveness of transfer learning, using pre-trained models from large datasets of adult human parsing, and data augmentation techniques to improve the accuracy and generalization capabilities of the model.

The hyperparameters of the models, including learning rate, batch size and data augmentation parameters, were optimized during the cross-validation process. The fusion of features extracted from both RGB and LWIR modalities significantly outperformed unimodal approaches, achieving a mean Intersection-over-Union of 0.87, a significant achievement given the limitations of the available datasets. Specifically, the fusion model achieved an IoU of 0.90 for the head, 0.88 for the arms, 0.86 for the legs and 0.69 for the torso, demonstrating the effectiveness of the model in accurately segmenting different parts of the body. This chapter has demonstrated the potential of multimodal deep learning for accurate and robust body part segmentation in infants, providing a powerful

tool for extracting accurate spatial temperature information from IRT images.

The next chapter, Chapter 4, moved from segmentation to extraction of temperature from a clinical study, validating the performance of the non-intrusive skin temperature measurement system on actual premature infants. This crucial step provided valuable insight into the practical feasibility and accuracy of the system in a clinical setting. The study design included data acquisition using both the IRT and RGB cameras to capture synchronized images of the infants. In addition, traditional contact-based temperature measurements were simultaneously acquired to provide a gold standard for comparison.

The study included 14 premature infants with gestational ages ranging from 27 to 31 weeks and birth weights ranging from 690 g to 1410 g. The IRT and RGB cameras were positioned on the infants and captured images at a frame rate of 1/s. To ensure accurate temperature measurements, a live temperature calibration method was used with a reference body placed inside the incubator. The temperature data extracted from the IRT images with the help of the segmentation model was compared both qualitatively and quantitatively with the contact-based temperature measurements. The study successfully demonstrated the feasibility of the IRT-based system, achieving good agreement with traditional contact-based temperature measurements. The system exhibited high spatial resolution, effectively detecting temperature variations in different regions of the body. In particular, it achieved a mean absolute error of 0.41 °C when comparing IRT measurements of the torso with contact-based measurements. This clinical validation provided strong evidence of the system's potential to improve neonatal care by providing non-invasive, spatially resolved temperature information, enhancing clinicians' ability to monitor and manage the thermal well-being of premature infants.

Chapter 5 introduced a novel neonatal phantom, a carefully constructed anatomical replica equipped with independent heating zones and temperature sensors, which served as a realistic simulation platform. This phantom, capable of mimicking the complex thermal behavior of premature infants, proved crucial in validating the temperature calibration algorithm and exploring different physiological and pathological scenarios, such as an increased central-peripheral temperature difference.

The phantom was designed using a 3D-printed base structure with carbon fiber heating elements and Pt1000 temperature sensors embedded in a silicone skin layer. This allowed independent temperature control and measurement in 13 different body zones. The thermal homogeneity of the phantom was optimized by incorporating boron nitride into the silicone skin layer to improve heat distribution. A series of tests were conducted to validate the phantom's ability to maintain stable temperature control, simulate different physiological scenarios and reproduce pathological conditions such as increased central-peripheral

temperature difference. The phantom successfully achieved these objectives, demonstrating its effectiveness as a valuable tool for testing and validating non-contact temperature measurement systems and control algorithms for neonatal incubators.

Finally, Chapter 6 explored the application of the non-contact temperature measurement system for skin servo control in neonatal incubators. A hardware-in-the-loop test bench was constructed, integrating the IRT system with the neonatal phantom. This setup allowed the controlled evaluation of skin servo control algorithms in a realistic environment simulating the dynamic thermal interactions between an infant and an incubator. Initial skin servo controllers, based on PI and gain-scheduled PI algorithms, were designed and evaluated for their ability to compensate for disturbances and variations in the infant's metabolism. Scenarios such as the opening of incubator hand ports and changes in metabolic heat production were simulated.

The control system was modeled and identified using Matlab/Simulink, incorporating first-order dynamics with uncertain gain and time constant to account for the inherent variability of the incubator and the thermal behavior of the infant. The controllers were then evaluated in both software-in-the-loop simulations and on the HIL test bench, comparing their performance in terms of mean square error and percentage of time within a defined setpoint range. The results demonstrated the feasibility of contactless skin servo control, achieving a mean square error of 0.079 for the PI controller and 0.088 for the gain-scheduled PI controller.

This research, which has led to the validation of a novel multimodal approach to contactless skin temperature management, provides a strong foundation for advancing the field. Future research efforts should focus on extending the clinical validation through larger-scale studies, investigating the impact of the contactless system on infant comfort, care practices and clinical outcomes.

Further optimization of the IRT system, including refinement of segmentation algorithms, development of advanced calibration methods and creation of more accurate infant thermal models, should be pursued to improve accuracy, robustness and usability. The development of more sophisticated and adaptive control algorithms, taking into account different physiological parameters and individual infant characteristics, could lead to personalized temperature management strategies that minimize variation and optimize infant thermoregulation.

In the future, incubators could be equipped with compact IRT systems inside the incubators that not only monitor but also adapt to the individual needs of each infant to prevent thermal stress. These systems present a significant opportunity compared to external cameras, minimizing interference with clinical care and eliminating errors from the infrared window. This technology

has immense potential to advance neonatal care by enabling earlier detection of thermal stress, facilitating timely intervention, improving thermoregulatory outcomes, enhancing patient comfort, reducing the risk of skin breakdown and infection, and reducing clinician workload. In addition, ethical considerations regarding data protection and informed consent must be carefully addressed in future development and implementation.

A Appendix

A.1 Backbone Testing

For investigating the backbones, a DenseNet-121 (7 million parameters) [HLVDMW17], a ResNet50 (23 million parameters) [HZRS16] and ResNext50 (25 million parameters) [XGD⁺17] were chosen, as they are commonly used backbones for segmentation tasks. Each backbone was integrated into the U-Net and trained on the Chennai dataset. The results of this evaluation are presented in Table A.1.

Tabelle A.1: IoUs per class for different backbones, training on the Chennai dataset.

Backbone	(Averaged) IoU				
	Head	Torso	Arms	Legs	Mean
DenseNet-121	0.89	0.37	0.79	0.89	0.73
ResNet50	0.88	0.38	0.78	0.86	0.72
ResNext50	0.89	0.38	0.77	0.85	0.72

A.2 Bode Plots

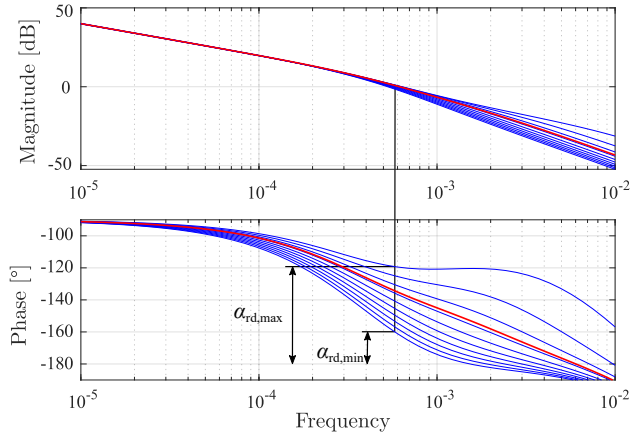


Abbildung A.1: Bode plot of the open loop with the gain scheduled PI controller ($K_p = 1$, uncertain $T_{Incubator}$). The nominal system is shown in red.

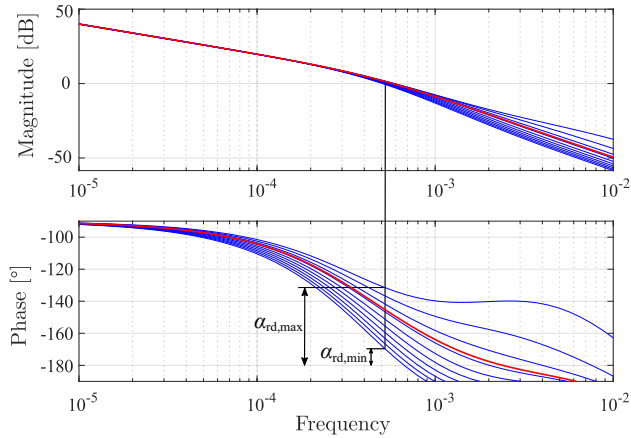


Abbildung A.2: Bode plot of the open loop with the gain scheduled PI controller ($K_p = 0.5$, uncertain $T_{Incubator}$). The nominal system is shown in red.

Peer-reviewed Journal Articles

1. **F. Voss***, L. Korn*, S. Leonhardt, and M. Walter: “Modeling of Flow-Dependent Blood Conductivity for Cardiac Bioimpedance”. *International Journal of Bioelectromagnetism*, 23(2), 2021.
2. S. Lyra*, **F. Voss***, A. Coenen, D. Blase, I. B. Aguirregomezcorta, D. U. Uguz, S. Leonhardt and C. H. Antink: “A Neonatal Phantom for Vital Signs Simulation”. *IEEE Transactions on Biomedical Circuits and Systems*, 15(5):949-959, 2021.
3. **F. Voss***, S. Lyra*, D. Blase, S. Leonhardt, and M. Lüken: “A Setup for Camera-Based Detection of Simulated Pathological States Using a Neonatal Phantom”. *Sensors*, 22(3):957, 2022.
4. **F. Voss**, N. Brechmann, S. Lyra, J. Rixen, S. Leonhardt, and C. Hoog Antink: “Multi-modal Body Part Segmentation of Infants Using Deep Learning”. *BioMedical Engineering OnLine*, 22(1):28, 2023.
5. **F. Voss**, P. Grünter, J. Wolski, S. Leonhardt, and M. Lüken: “Hardware-in-the-Loop Setup for Contactless Skin Servo Control of Neonatal Incubators”. *Biomedical Signal Processing and Control*, 88:105628, 2024.
6. **F. Voss**, S. Lyra, C. Nöhl, M. Brasche, K. Heimann, L. Hensel, T. Orlikowsky, S. Leonhardt, and M. Lueken: “Spatial Temperature Monitoring of premature infants using a multi-modal camera system”. *BioMedical Engineering OnLine*, Submitted and under review. .

*Shared first authorship

Conference Proceedings

1. **F. Voss**, C. Stapper, S. Lyra: “Multi-Modal Body Part Detection of Infants using Deep Learning”. *26th International Student Conference on Electrical Engineering (POSTER 2022)*. Prague, Czech Republic, May 12, 2022.
2. **F. Voss**, J. Wolski, D. Blase, S. Leonhardt and M. Lüken: “Voss, Florian, et al. "Live Temperature Calibration for Neonatal Thermography". *World Congress on Medical Physics and Biomedical Engineering (IU-PESM WC2022)*. Singapore, May 12-17, 2022.

3. **F. Voss**, L. Tiltmann, S. Lyra, S. Leonhardt, M. Lüken: “A Physical Phantom for the Simulation of Neonatal Thermoregulation”. *45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. Sydney, Australia, July 24-27, 2023.
4. **F. Voss**, M. Lüken, S. Leonhardt: “Regression-based Correction for Infrared Thermography in Neonatal Incubators”. *57th Annual Meeting of the German Society of Biomedical Engineering(BMT 2023)*. Duisburg, Germany, September 26-28, 2023.
5. **F. Voss**, K. Koessler, S. Leonhardt, M. Lüken: “Camera Fusion for Improving Body Part Segmentation of Preterm Infants”. *Workshop Biosignals*. Göttingen, Germany, February 28 - March 1, 2024.
6. **F. Voss**, S. Leonhardt, M. Lüken: “Skin Servo Control for Neonatal Incubators: A Novel Approach using Infrared Thermography”. *17th Interdisciplinary Symposium Automed*. Villingen-Schwenningen, Germany, September 11-13, 2024.

Supervised Student Theses

1. A. Coenen: *Development and construction of a neonate phantom to reproduce physiological processes*. Master's thesis, 2020.
2. B. Fathy: *Segmentation of neonates in thermal images using Deep Learning*. Bachelor's thesis, 2021.
3. H. Breidenich: *Development of a body pose estimator for thermal images using deep learning*. Bachelor's thesis, 2021.
4. C. Stapper: *Robust extraction of the temperature distribution of neonates using Deep Learning*. Master's thesis, 2021.
5. N. Brechmann: *Robust skin detection of neonates using image fusion and deep learning*. Master's thesis, 2021.
6. D. Blase: *Camera-based detection of pathology patterns using a neonatal phantom and Deep Learning*. Master's thesis, 2021.
7. J. Wolski: *Smart Temperature Assessment for Neonatal Thermography*. Bachelor's thesis, 2021.
8. L. Tiltmann: *Development of a thermal phantom*. Bachelor's thesis, 2022.
9. P. Grünter: *Development of a Hardware-in-the-Loop Test Bench for a Neonatal Incubator*. Bachelor's thesis, 2022.
10. J. Jordan: *Automated Disturbance Detection in Neonatal Incubators using Thermal Imaging*. Bachelor's thesis, 2023.
11. S. Rütten: *Development of a miniaturised Thermography System for Monitoring Preterm Infants Inside an Incubator*. Bachelor's thesis, 2023.
12. S. Tang: *Modeling a Neonatal Incubator for Optimal Temperature Control*. Bachelor's thesis, 2023.
13. K. Kösseler: *Camera-based Temperature Extraction of Premature Infants using Deep Learning*. Master's thesis, 2023.
14. T. Meyer: *Development of a Deep-Learning based Multimodal Camera System for Thermal Monitoring of Premature Infants*. Master's thesis, 2024.
15. C. Nöhl: *Spatial Temperature Monitoring of Preterm Infants using Deep Learning*. Master's thesis, 2024.

Literaturverzeichnis

- [Adm21] U. S. F. . D. Administration: “Thermal imaging systems (infrared thermographic systems / thermal imaging cameras)”. <https://www.fda.gov/medical-devices/general-hospital-devices-and-supplies/thermal-imaging-systems-infrared-thermographic-systems-thermal-imaging-cameras>, 2021.
- [AEHT⁺99] A. Apedoh, A. El Hajajji, F. Telliez, B. Bouferrache, J.-P. Libert, and A. Rachid: “Mannequin-assessed dry-heat exchanges in the incubator-nursed newborn.” *Biomedical Instrumentation & Technology*, 33(5):446–454, 1999.
- [AFP⁺20] C. H. Antink, J. C. M. Ferreira, M. Paul, S. Lyra, K. Heimann, S. Karthik, J. Joseph, K. Jayaraman, T. Orlikowsky, M. Sivaprakasam, and S. Leonhardt: “Fast body part segmentation and tracking of neonatal video data using deep learning”. *Medical & Biological Engineering & Computing*, 58(12):3049–3061, 2020.
- [AHB⁺12] A. K. Abbas, K. Heimann, V. Blazek, T. Orlikowsky, and S. Leonhardt: “Neonatal infrared thermography imaging: analysis of heat flux during different clinical scenarios”. *Infrared Physics & Technology*, 55(6):538–548, 2012.
- [AHH⁺22] H. Asano, E. Hirakawa, H. Hayashi, K. Hamada, Y. Asayama, M. Oohashi, A. Uchiyama, and T. Higashino: “A method for improving semantic segmentation using thermographic images in infants”. *BMC Medical Imaging*, 22(1):1–13, 2022.
- [AHJ⁺11] A. K. Abbas, K. Heimann, K. Jergus, T. Orlikowsky, and S. Leonhardt: “Neonatal non-contact respiratory monitoring based on real-time infrared thermography”. *Biomedical engineering online*, 10:1–17, 2011.
- [Ahn10] Y. Ahn: “Formula and scale for body surface area estimation in high-risk infants”. *Collegium antropologicum*, 34(4):1273–1280, 2010.
- [Bac05] M. Bacic: “On hardware-in-the-loop simulation”. *Proceedings of the 44th IEEE Conference on Decision and Control*, 3194–3198. IEEE, 2005.

- [Bar63] R. B. Barnes: “Thermography of the human body: Infrared-radiant energy provides new concepts and instrumentation for medical diagnosis”. *Science*, 140(3569):870–877, 1963.
- [Bau08] S. Baumgart: “Iatrogenic hyperthermia and hypothermia in the neonate”. *Clinics in perinatology*, 35(1):183–197, 2008.
- [BBCL17] O. Bonner, K. Beardsall, N. Crilly, and J. Lasenby: “There were more wires than him: the potential for wireless patient monitoring in neonatal intensive care”. *BMJ Innovations*, 3(1), 2017. Publisher: BMJ Specialist Journals Section: Health IT, systems and process innovations.
- [BDL⁺01] S. Blackburn, D. DePaul, L. Loan, K. Marbut, L. Taquino, K. Thomas, and S. Wilson: “Neonatal thermal care, part ii: microbial growth under temperature probe covers”. *Neonatal Network*, 20(3):19–23, 2001.
- [Bie18] A. Bielecki: *Models of Neurons and Perceptrons, Studies in Computational Intelligence Ser*, vol. v. 770. Springer International Publishing AG, 2018.
- [BMR⁺20] M. Bruckner, L. P. Mileder, A. Richter, N. Baik-Schneditz, B. Schwabegger, C. Binder-Heschl, B. Urlesberger, and G. Pichler: “Association between regional tissue oxygenation and body temperature in term and preterm infants born by caesarean section”. *Children*, 7(11):205, 2020.
- [BTE⁺06] K. Belghazi, P. Tourneux, E. B. Elabbassi, L. Ghyselen, S. Delanaud, and J.-P. Libert: “Effect of posture on the thermal efficiency of a plastic bag wrapping in neonate: assessment using a thermal sweating mannequin”. *Medical physics*, 33(3):637–644, 2006.
- [Buc80] G. Buchsbaum: “A spatial processor model for object colour perception”. *Journal of the Franklin institute*, 310(1):1–26, 1980.
- [Bud07] P. Budin: *The Nursling: the feeding and hygiene of premature & full-term infants*. Caxton Publishing Company, 1907.
- [Cho18] F. Chollet: *Deep learning with Python*. Manning Publications Co, Shelter Island, NY, 2018.

- [CML⁺14] X. Chen, R. Mottaghi, X. Liu, S. Fidler, R. Urtasun, and A. Yuille: “Detect what you can: Detecting and representing objects using holistic models and body parts”. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1971–1978, 2014.
- [CVJ⁺17] S. Chaichulee, M. Villarroel, J. Jorge, C. Arteta, G. Green, K. McCormick, A. Zisserman, and L. Tarassenko: “Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring”. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, 266–272. IEEE, Washington DC, 2017.
- [CWH⁺05] K. Cramer, N. Wiebe, L. Hartling, E. Crumley, and S. Vohra: “Heat loss prevention: a systematic review of occlusive skin wrap for premature neonates”. *Journal of Perinatology*, 25(12):763–769, 2005.
- [DBN⁺18] Y. S. Dossso, A. Bekele, S. Nizami, C. Aubertin, K. Greenwood, J. Harrold, and J. R. Green: “Segmentation of patient images in the neonatal intensive care unit”. *2018 IEEE Life Sciences Conference (LSC)*, 45–48. IEEE, Piscataway, NJ, 2018.
- [DDP⁺12] C. Deguines, P. Decima, A. Pelletier, L. Degrugilliers, L. Ghyselen, and P. Tourneux: “Variations in incubator temperature and humidity management: a survey of current practice”. *Acta Paediatrica*, 101(3):230–235, 2012.
- [DLRR⁺85] D. A. Ducker, A. J. Lyon, R. Ross Russell, C. A. Bass, and N. McIntosh: “Incubator temperature control: effects on the very low birthweight infant”. *Archives of Disease in Childhood*, 60(10):902, 1985.
- [Dra19] Draegerwerk AG & Co. KGaA: “Heat control device for an incubator for infants, incubator and method for the heat control of an incubator”. US10314756B2, 2019.
- [DT17] T. DeVries and G. W. Taylor: “Improved regularization of convolutional neural networks with cutout”. *arXiv preprint arXiv:1708.04552*, 2017.

- [DZM⁺21] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun: “Repvgg: Making vgg-style convnets great again”. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 13733–13742, 2021.
- [EBM⁺01] E. B. Elabbassi, V. Bach, M. Makki, S. Delanaud, F. Telliez, A. Leke, and J.-P. Libert: “Assessment of dry heat exchanges in newborns: influence of body position and clothing in sids”. *Journal of Applied Physiology*, 91(1):51–56, 2001.
- [Eur11] European Foundation for the Care of Newborn Infants (EFCNI): “Caring for tomorrow”, 2011.
URL https://www.efcni.org/wp-content/uploads/2018/03/EFCNI_WP_01-26-12FIN.pdf
- [Eur19a] European Foundation for the Care of Newborn Infants (EFCNI): “Key facts preterm birth”, 2019.
URL <https://www.efcni.org/health-topics/key-facts-preterm-birth/>
- [Eur19b] European Foundation for the Care of Newborn Infants (EFCNI): “Preterm birth”, 2019.
URL https://www.efcni.org/wp-content/uploads/2019/10/2019_10_21_Factsheet_Preterm_Birth_web.pdf
- [EvW⁺10] M. Everingham, L. van Gool, C. K. I. Williams, J. Winn, and A. Zisserman: “The pascal visual object classes (voc) challenge”. *International Journal of Computer Vision*, 88(2):303–338, 2010.
- [FBNK98] R. T. Frankenberger, O. Bussmann, W. Nahm, and E. Konecny: “Simulation model of heat loss in preterm infants”. *Biomedizinische Technik*, 43(5):137–143, 1998.
- [FH20] F. Farahnakian and J. Heikkonen: “Deep learning based multi-modal fusion architectures for maritime vessel detection”. *Remote Sensing*, 12(16):2509, 2020.
- [FK13] T. R. Fenton and J. H. Kim: “A systematic review and meta-analysis to revise the fenton growth chart for preterm infants”. *BMC pediatrics*, 13(1):1–13, 2013.

- [FMGW22] O. Frigo, L. Martin-Gaffe, and C. Wacongne: “Doodlenet: Double deeplab enhanced feature fusion for thermal-color semantic segmentation”. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CV-PRW)*, 3020–3028, 2022.
- [GLL⁺18] K. Gong, X. Liang, Y. Li, Y. Chen, M. Yang, and L. Lin: “Instance-level human parsing via part grouping network”. *Proceedings of the European Conference on Computer Vision (ECCV)*, 770–785, 2018.
- [GPK13] D.-W. Gu, P. H. Petkov, and M. M. Konstantinov: *Robust Control Design with MATLAB® (second edition)*, 2013.
- [GT19] A. I. Georgevici and M. Terblanche: “Neural networks and deep learning: a brief introduction”. *Intensive care medicine*, 45(5):712–714, 2019.
- [Han21] A. Hannouch: *Numerical and Experimental Analyses of the Heat Transfer inside Infant Incubators using 3D Printed Thermal Manikin*. Ph.D. thesis, Université d’Angers, 2021.
- [Hea22] G. Healthcare: “Giraffe carestation neonodevelopmentally supportive, family-centered critical care environment for newborns”. <https://www.gehealthcare.de/products/perinatal-care/giraffe-omnibed-carestation>, Online; accessed 30.11.2022.
- [Hen24] Henze Boron Nitride Products AG: “Hebofill bornitrid-pulverqualitäten”. <https://www.henze-bnp.de/hebofill-bornitrid-pulver.php>, accessed 30.01.2024.
- [HHA⁺22a] K. Hamada, E. Hirakawa, H. Asano, H. Hayashi, T. Mine, T. Ichikawa, and Y. Nagata: “Infrared thermography with high accuracy in a neonatal incubator”. *Annals of Biomedical Engineering*, 50(5):529–539, 2022.
- [HHA⁺22b] K. Hamada, E. Hirakawa, H. Asano, H. Hayashi, T. Mine, T. Ichikawa, and Y. Nagata: “Infrared thermography with high accuracy in a neonatal incubator”. *Annals of Biomedical Engineering*, 50(5):529–539, 2022.

- [HKV19] F. Hutter, L. Kotthoff, and J. Vanschoren: *Automated machine learning: methods, systems, challenges*. Springer Nature, 2019.
- [HLVDMW17] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger: “Densely connected convolutional networks”. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4700–4708, 2017.
- [HMS21] E. Hering, R. Martin, and M. Stohrer: *Physik für Ingenieure*. Springer Berlin Heidelberg, 13th ed., 2021.
- [Hol00] G. C. Holst: *Common sense approach to thermal imaging*. SPIE Optical Engineering Press, 2000.
- [HWK⁺17] Q. Ha, K. Watanabe, T. Karasawa, Y. Ushiku, and T. Harada: “Mfnet: Towards real-time semantic segmentation for autonomous vehicles with multi-spectral scenes”. *IROS Vancouver 2017*, 5108–5115. IEEE, Piscataway, NJ, 2017.
- [HWW⁺19] Z. Huang, X. Wang, Y. Wei, L. Huang, H. Shi, W. Liu, and T. S. Huang: “Ccnet: Criss-cross attention for semantic segmentation”. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 603–612, 2019.
- [HZRS16] K. He, X. Zhang, S. Ren, and J. Sun: “Deep residual learning for image recognition”. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778, 2016.
- [IAFS17] I. Ilievski, T. Akhtar, J. Feng, and C. Shoemaker: “Efficient hyperparameter optimization for deep learning algorithms using deterministic rbf surrogates”. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1), 2017.
- [Inf23] InfraTec GmbH: “Infrared radiation”, 2023.
URL <https://www.infratec-infrared.com/sensor-division/service-support/glossary/infrared-radiation/>
- [Int18] Intertronics: *TECHNICAL BULLETIN Thermally Conductive Silicones*, 2018.

- [IS15] S. Ioffe and C. Szegedy: “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. *International conference on machine learning*, 448–456, 2015.
- [ISS99] R. Isermann, J. Schaffnit, and S. Sinsel: “Hardware-in-the-loop simulation for the design and testing of engine-control systems”. *Control Engineering Practice*, 7(5):643–653, 1999.
- [JM13] H. Javed and A. Mahmood: “A study of thermodynamics and bond graph modelling of evaporation in infant incubator”. *2013 IEEE 9th International Conference on Emerging Technologies (ICET)*, 1–6. IEEE, 2013.
- [Jol61] H. Jolly: “A controlled study of the effect of temperature on premature babies”. *Proceedings of the Royal Society of Medicine-London*, 54(9):742–742, 1961.
- [JPDE⁺17] K. Jost, I. Pramana, E. Delgado-Eckert, N. Kumar, A. N. Datta, U. Frey, and S. M. Schulzke: “Dynamics and complexity of body temperature in preterm infants nursed in incubators”. *PLoS One*, 12(4):e0176670, 2017.
- [KB14] D. P. Kingma and J. Ba: “Adam: A method for stochastic optimization”. *International Conference on Learning Representations*, 2014.
- [KD14] R. B. Knobel-Dail: “Role of effective thermoregulation in premature neonates”. *Research and Reports in Neonatology*, 4:147–156, 2014.
- [KDHDS⁺17] R. B. Knobel-Dail, D. Holditch-Davis, R. Sloane, B. Guenther, and L. M. Katz: “Body temperature in premature infants during the first week of life: Exploration using infrared thermal imaging”. *Journal of thermal biology*, 69:118–123, 2017.
- [KGR11] R. B. Knobel, B. D. Guenther, and H. E. Rice: “Thermoregulation and thermography in neonatal physiology and disease”. *Biological research for nursing*, 13(3):274–282, 2011.
- [KHDSW09] R. B. Knobel, D. Holditch-Davis, T. A. Schwartz, and J. Wimmer: “Extremely low birth weight preterm infants

- lack vasomotor response in relationship to cold body temperatures at birth”. *Journal of Perinatology*, 29(12):814–821, 2009.
- [Kim19] Kim: “Little baby 3d-modell”, 20-11-2019. (Last accessed 28-11-2022).
URL <https://free3d.com/3d-model/little-baby-56644.html>
- [KOA10] S. Kemaloglu, G. Ozkoc, and A. Aytac: “Properties of thermally conductive micro and nano size boron nitride reinforced silicon rubber composites”. *Thermochimica Acta*, 499(1-2):40–47, 2010.
- [Koc16] J. Koch: “Thermoregulation des Menschen”. S. Leonhardt and M. Walter (eds.), *Medizintechnische Systeme: Physiologische Grundlagen, Gerätetechnik und automatisierte Therapieführung*, 285–297. Springer, Berlin, Heidelberg, 2016.
- [KSK19] I. Krucińska, E. Skrzetuska, and K. Kowalski: “Application of a thermal mannequin to the assessment of the heat insulating power of protective garments for premature babies”. *Autex Research Journal*, 19(2):134–146, 2019.
- [Kun14] L. I. Kuncheva: *Combining pattern classifiers: methods and algorithms*. John Wiley & Sons, Hoboken, 2014.
- [KVDBVDV⁺14] J. H. Klaessens, M. Van Den Born, A. Van Der Veen, J. Sikkens-Van De Kraats, F. A. van den Dungen, and R. M. Verdaasdonk: “Development of a baby friendly non-contact method for measuring vital signs: first results of clinical measurements in an open incubator at a neonatal intensive care unit”. *Advanced biomedical and clinical diagnostic systems XII*, vol. 8935, 257–263. SPIE, 2014.
- [LBJH13] K. Lunze, D. E. Bloom, D. T. Jamison, and D. H. Hamer: “The global burden of neonatal hypothermia: systematic review of a major challenge for newborn survival”. *BMC medicine*, 11(1):1–11, 2013.
- [LHKS19] J. H. Lee, M.-K. Han, D. W. Ko, and I. H. Suh: “From big to small: Multi-scale local planar guidance for monocular depth estimation”. *arXiv preprint arXiv:1907.10326*, 2019.

- [Lju99] L. Ljung: *System identification: theory for the user (second edition)*. Prentice Hall, 1999.
- [LOH⁺16] L. Liu, S. Oza, D. Hogan, Y. Chu, J. Perin, J. Zhu, J. E. Lawn, S. Cousens, C. Mathers, and R. E. Black: “Global, regional, and national causes of under-5 mortality in 2000–15: an updated systematic analysis with implications for the sustainable development goals”. *The Lancet*, 388(10063):3027–3035, 2016.
- [LP15] A. Lyon and P. Püscher: “ThermoMonitoring: A step forward in neonatal intensive care”, 2015.
URL <https://www.draeger.com/Content/Documents/Content/thermomonitoring-bk-9097384-us.pdf>
- [LRH⁺22] S. Lyra, J. Rixen, K. Heimann, S. Karthik, J. Josphe, K. Jayaraman, T. Orlikowsky, S. Sivaprakasam, S. Leonhardt, and C. Hoog Antink: “Camera fusion for real-time temperature monitoring of neonates using deep learning”. *Medical & Biological Engineering & Computing*, 2022. (accepted for publication).
- [LSB⁺07] A. R. Laptook, W. Salhab, B. Bhaskar, *et al.*: “Admission temperature of low birth weight infants: predictors and associated morbidities”. *Pediatrics*, 119(3):e643–e649, 2007.
- [LSC19] A. Lubkowska, S. Szymański, and M. Chudecka: “Surface body temperature of full-term healthy newborns immediately after birth - pilot study”. *International journal of environmental research and public health*, 16(8):1312, 2019.
- [LSD15] J. Long, E. Shelhamer, and T. Darrell: “Fully convolutional networks for semantic segmentation”. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3431–3440, 2015.
- [LSM⁺21] I. Lorato, S. Stuijk, M. Meftah, D. Kommers, P. Andriessen, C. van Pul, and G. de Haan: “Towards continuous camera-based respiration monitoring in infants”. *Sensors*, 21(7):2268, 2021.
- [LSX⁺15] Y. Lyu, P. S. Shah, Y. Y. Xiang, R. Warre, B. Piedboeuf, A. Deshpandey, M. Dunn, S. K. Lee, C. N. Network, *et al.*:

- “Association between admission temperature and mortality and major morbidity in preterm infants born at fewer than 33 weeks of gestation”. *JAMA pediatrics*, 169(4):e150277–e150277, 2015.
- [LVC⁺21] S. Lyra, F. Voss, A. Coenen, D. Blase, I. B. Aguirregomezcor-ta, D. U. Uguz, S. Leonhardt, and C. H. Antink: “A neonatal phantom for vital signs simulation”. *IEEE Transactions on Biomedical Circuits and Systems*, 15(5):949–959, 2021.
- [LW16] S. Leonhardt and M. Walter: *Medizintechnische Systeme: Physiologische Grundlagen, Gerätetechnik und automatisier-te Therapieführung*. Springer Berlin Heidelberg, 2016.
- [LW19] H. Li and X.-J. Wu: “Densefuse: A fusion approach to in-ffrared and visible images”. *IEEE Transactions on Image Processing*, 28(5):2614–2623, 2019.
- [LWL⁺20] K. Lin, L. Wang, K. Luo, Y. Chen, Z. Liu, and M.-T. Sun: “Cross-domain complementary learning using pose for multi-person part segmentation”. *IEEE Transactions on Circuits and Systems for Video Technology*, 1066–1078, 2020.
- [LXWY20] P. Li, Y. Xu, Y. Wei, and Y. Yang: “Self-correction for hu-man parsing”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.
- [Lyo07] A. Lyon: “Temperature control in the neonate”. *Paediatrics and Child Health*, 18(4):155–160, 2007.
- [MC⁺03] M. McCormick, P. Cooper, *et al.*: *Managing newborn pro-blems: a guide for doctors, nurses, and midwives*. World Health Organization, 2003.
- [Mic19] Michael Zemlin (Homburg; federführend), Angelika Berger (Wien), Axel Franz (Tübingen), Christian Gille (Tübingen), Christoph Härtel (Lübeck), Helmut Küster (Göttingen), An-dreas Müller (Bonn), Frank Pohlandt (Ulm), Arne Simon (Homburg) und Waltraud Merz (Bonn): “Leitlinie der ge-sellschaft für neonatologie und pädiatrische intensivmedi-zin, der deutschen gesellschaft für kinder- und jugendme-dizin, der gesellschaft für pädiatrische gastroenterologie und

ernährung und der deutschen gesellschaft für kinderchirurgie”. *AWMF online, AWMF-Leitlinien-Register Nr. 024/008*, Entwicklungsstufe 2k, 2019.

- [MVBDB08] P. H. Muller, L. Van Berkel, and A. De Beaufort: “Axillary and rectal temperature measurements poorly agree in newborn infants”. *Neonatology*, 94(1):31–34, 2008.
- [NBAR04] S. Nabi, M. Balike, J. Allen, and K. Rzemien: “An overview of hardware-in-the-loop testing systems at visteon”. *SAE Technical paper*, 2004.
- [ODR15] T. Oranges, V. Dini, and M. Romanelli: “Skin physiology of the neonate and infant: clinical implications”. *Advances in wound care*, 4(10):587–595, 2015.
- [ORS⁺16] Z. Ostrowski, M. Rojczyk, I. Szczygiel, J. Łaszczyk, and A. J. Nowak: “Dry heat loses of newborn baby in infant care bed: use of a thermal manikin”. *Journal of Physics: Conference Series*, 745:032087, 2016.
- [OVB⁺16] G. L. Oliveira, A. Valada, C. Bollen, W. Burgard, and T. Brox: “Deep learning for human part discovery in images”. *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 1634–1641, 2016.
- [PW17] L. Perez and J. Wang: “The effectiveness of data augmentation in image classification using deep learning”. *Convolutional Neural Networks Vis. Recognit*, 11, 2017.
- [PYG⁺18] C. B. Pereira, X. Yu, T. Goos, I. Reiss, T. Orlikowsky, K. Heimann, B. Venema, V. Blazek, S. Leonhardt, and D. Teichmann: “Noncontact monitoring of respiratory rate in newborn infants using thermal imaging”. *IEEE transactions on Biomedical Engineering*, 66(4):1105–1114, 2018.
- [RDS⁺15] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei: “Imagenet large scale visual recognition challenge”. *International Journal of Computer Vision*, 115(3):211–252, 2015.

- [RFB15] O. Ronneberger, P. Fischer, and T. Brox: “U-net: Convolutional networks for biomedical image segmentation”. *International Conference on Medical image computing and computer-assisted intervention*, 234–241, 2015.
- [RLH⁺19] T. Ruan, T. Liu, Z. Huang, Y. Wei, S. Wei, Y. Zhao, and T. Huang: “Devil in the details: Towards accurate single and multiple human parsing”. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 4814–4821, 2019.
- [RMH09] N. P. Reddy, G. Mathur, and S. I. Hariharan: “Toward a fuzzy logic control of the infant incubator”. *Annals of Biomedical Engineering*, 37(10):2146–2152, 2009.
- [Rog12] A. Rogalski: “Progress in focal plane array technologies”. *Progress in Quantum Electronics*, 36(2-3):342–473, 2012.
- [SAU13] J. Smith, G. Alcock, and K. Usher: “Temperature measurement in the preterm and term neonate: a review of the literature”. *Neonatal Network*, 32(1):16–25, 2013.
- [SBHT92] I. Sarman, D. Bolin, I. Holmér, and R. Tunell: “Assessment of thermal conditions in neonatal care: use of a manikin of premature baby size”. *American journal of perinatology*, 9(04):239–246, 1992.
- [SBP09] W. Siems, A. Bremer, and J. Przyklenk: *Allgemeine Krankheitslehre für Physiotherapeuten*. Springer-Verlag, Berlin-Heidelberg, 2009.
- [Sch04] Schuster: *Infrarotthermographie*. John Wiley & Sons, Inc, 2004.
- [Sed95] G. Sedin: “Neonatal Heat Transfer, Routes of Heat Loss and Heat Gain”. A. Okken and J. Koch (eds.), *Thermoregulation of Sick and Low Birth Weight Neonates*, 21–36. Springer, Berlin, Heidelberg, 1995.
- [SFB58] W. A. Silverman, J. W. Fertig, and A. P. Berger: “The influence of the thermal environment upon the survival of newly born premature infants”. *Pediatrics*, 22(5):876–886, 1958.
- [SG13] C. P. Speer and M. Gahr: *Pädiatrie*. Springer, Berlin, Heidelberg, Berlin, Heidelberg, 2013.

- [Sin02] J. C. Sinclair: “Servo-control for maintaining abdominal skin temperature at 36c in low birth weight infants”. *Cochrane Database of Systematic Reviews*, (1), 2002.
- [SK19] C. Shorten and T. M. Khoshgoftaar: “A survey on image data augmentation for deep learning”. *Journal of Big Data*, 6(1), 2019.
- [Smi18] L. N. Smith: “A disciplined approach to neural network hyper-parameters: Part 1–learning rate, batch size, momentum, and weight decay”. *US Naval Research Laboratory Technical Report*, 2018.
- [SMTK⁺08] C. K. Shapiro-Mendoza, K. M. Tomashek, M. Kotelchuck, W. Barfield, A. Nannini, J. Weiss, and E. Declercq: “Effect of late-preterm birth and maternal medical conditions on newborn morbidity risk”. *Pediatrics*, 121(2):e223–e232, 2008.
- [SSM⁺17] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. Kruthiventi, and R. V. Babu: “An introduction to deep convolutional neural nets for computer vision”. *Deep Learning for Medical Image Analysis*, 25–52. Elsevier, 2017.
- [Suk17] H.-I. Suk: “An introduction to neural networks and deep learning”. *Deep Learning for Medical Image Analysis*, 3–24. Elsevier, 2017.
- [SVP⁺20] A. Somboonkaew, S. Vuttivong, P. Prempre, R. Amarit, S. Chanhorm, K. Chaitavon, S. Porntheeraphat, and S. Sumriddetchkajorn: “Temperature-compensated infrared-based low-cost mobile platform module for mass human temperature screening”. *Applied optics*, 59(17):E112–E117, 2020.
- [SW98] J. R. Speakman and S. Ward: “Infrared thermography: principles and applications”. *Zoology-Jena-*, 101:224–232, 1998.
- [Sze11] R. Szeliski: *Computer Vision*. Springer London, 2011.
- [SZL19] Y. Sun, W. Zuo, and M. Liu: “Rtfnnet: Rgb-thermal fusion network for semantic segmentation of urban scenes”. *IEEE Robotics and Automation Letters*, 4(3):2576–2583, 2019.

- [TB99] K. A. Thomas and R. Burr: “Preterm infant thermal care: differing thermal environments produced by air versus skin servo-control incubators”. *Journal of Perinatology*, 19(4):264–70, 1999.
- [Tho03] K. A. Thomas: “Preterm infant thermal responses to caregiving differ by incubator control mode”. *Journal of Perinatology*, 23(8):640–645, 2003.
- [TSG⁺16] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and L. Jianming: “Convolutional neural networks for medical image analysis: Full training or fine tuning?” *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, 2016.
- [Unb08] H. Unbehauen: *Regelungstechnik I: Klassische Verfahren zur Analyse und Synthese linearer kontinuierlicher Regelsysteme, Fuzzy-Regelsysteme*, vol. 15. Vieweg+Teubner, Wiesbaden, 2008.
- [Uni21] United Nations Statistical Commission and Inter-Agency and Expert Group on SDG Indicators (IAEG-SDGs): “Global indicator framework for the sustainable development goals and targets of the 2030 agenda for sustainable development”. <https://unstats.un.org/sdgs/indicators/indicators-list/>, 2017-2021.
- [UUW⁺20] UNICEF, United Nations Children’s Fund, WHO and World Bank Group, UN Population Division, and UN IGME: “Levels and trends in child mortality”. *March of Dimes, PM-NCH, Save the Children*, 2020.
- [UVG⁺14] R. Usamentiaga, P. Venegas, J. Guerediaga, L. Vega, J. Mollada, and F. G. Bulnes: “Infrared thermography for temperature measurement and non-destructive testing”. *Sensors*, 14(7):12305–12348, 2014.
- [VBL⁺23] F. Voss, N. Brechmann, S. Lyra, J. Rixen, S. Leonhardt, and C. Hoog Antink: “Multi-modal body part segmentation of infants using deep learning”. *BioMedical Engineering OnLine*, 22(1):28, 2023.

- [VCJ⁺19] M. Villarroel, S. Chaichulee, J. Jorge, S. Davis, G. Green, C. Arteta, A. Zisserman, K. McCormick, P. Watkinson, and L. Tarassenko: “Non-contact physiological monitoring of pre-term infants in the neonatal intensive care unit”. *NPJ digital medicine*, 2:128, 2019.
- [VDDP18] A. Voulodimos, N. Doulamis, A. Doulamis, and E. Protopadakis: “Deep learning for computer vision: A brief review”. *Computational intelligence and neuroscience*, 2018, 2018.
- [VGW⁺24] F. Voss, P. Grünter, J. Wolski, S. Leonhardt, and M. Lüken: “Hardware-in-the-loop setup for contactless skin servo control of neonatal incubators”. *Biomedical Signal Processing and Control*, 88:105628, 2024.
- [VKLL24] F. Voss, K. K. Koessler, S. Leonhardt, and M. Lüken: “Camera fusion for improving body part segmentation of pre-term infants”. Workshop Biosignals, Göttingen, 2024.
- [VLB⁺22] F. Voss, S. Lyra, D. Blase, S. Leonhardt, and M. Lüken: “A setup for camera-based detection of simulated pathological states using a neonatal phantom”. *Sensors*, 22(3):957, 2022.
- [VLL24] F. Voss, S. Leonhardt, and M. Lueken: “Skin servo control for neonatal incubators: A novel approach using infrared thermography”. *Proceedings on Automation in Medical Engineering*, 1(1):1–1, 2024.
- [VLN⁺25] F. Voss, S. Lyra, C. Nöhl, M. Brasche, K. Heimann, L. Hensel, T. Orlikowsky, S. Leonhardt, and M. Lueken: “Spatial temperature monitoring of preterm infants using a multi-modal camera setup”. *Submitted to Medical & Biological Engineering & Computing*, 2025.
- [VS86] A. W. Van Herwaarden and P. M. Sarro: “Thermal sensors based on the seebeck effect”. *Sensors and Actuators*, 10(3-4):321–346, 1986.
- [VTL⁺23] F. Voss, L. Tiltmann, S. Lyra, S. Leonhardt, and M. Lueken: “A physical phantom for the simulation of neonatal thermoregulation”. *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1–4. IEEE, 2023.

- [VWB⁺22] F. Voss, J. Wolski, D. Blase, S. Leonhardt, and M. Lüken: “Live temperature calibration for neonatal thermography”. World Congress on Medical Physics and Biomedical Engineering, Singapur, 2022.
- [Wey23] Weyer GmbH: “Thermocare vita”. <https://www.weyermed.com/produkte/produktuebersicht/thermocare-vita>, accessed 17/11/2023.
- [WM07] S. Waldron and R. MacKinnon: “Neonatal thermoregulation”. *Infant*, 3(3):101–104, 2007.
- [Wor97] World Health Organization: “Maternal. Newborn Health. Thermal Protection of the Newborn: A Practical Guide”, 1997.
URL <https://apps.who.int/iris/handle/10665/63986>
- [Wor12] World Health Organization: “Born too soon: The global action report on preterm birth”. *March of Dimes, PMNCH, Save the Children*, 2012.
- [Wor18] World Health Organization: “Preterm birth”, 19th February 2018.
URL <https://www.who.int/news-room/fact-sheets/detail/preterm-birth>
- [Wor20] World Health Organization: “Newborns: improving survival and well-being”, 19 September 2020.
URL <https://www.who.int/news-room/fact-sheets/detail/newborns-reducing-mortality>
- [WZA⁺05] Z. Wang, D. Ziou, C. Armenakis, D. Li, and Q. Li: “A comparative analysis of image fusion methods”. *IEEE Transactions on Geoscience and Remote Sensing*, 43(6):1391–1402, 2005.
- [XGD⁺17] S. Xie, R. Girshick, P. Dollár, Z. Tu, and K. He: “Aggregated residual transformations for deep neural networks”. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1492–1500, 2017.
- [YSWJ19] L. Yang, Q. Song, Z. Wang, and M. Jiang: “Parsing r-cnn for instance-level human analysis”. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 364–373, 2019.

- [YZZ⁺18] B. Yang, S.-H. Zhang, Y.-F. Zou, W.-S. Ma, G.-J. Huang, and M.-D. Li: “Improving the thermal conductivity and mechanical properties of two-component room temperature vulcanized silicone rubber by filling with hydrophobically modified SiO₂-graphene nanohybrids”. *Chinese Journal of Polymer Science*, 37(2):189–196, 2018.
- [ZAB⁺18] J. Zeitlin, S. Alexander, H. Barros, B. Blondel, M. Delnord, M. Durox, M. Gissler, A. Macfarlane, A. Mohangoo, *et al.*: “European perinatal health report. core indicators of the health and care of pregnant women and babies in europe in 2015. world health organization”, 2018.
- [ZCVZ15] R. Zhang, S. A. Candra, K. Vetter, and A. Zakhor: “Sensor fusion for semantic segmentation of urban scenes”. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 1850–1857. IEEE, Seattle, 2015.
- [ZD] K. Zhou and J. C. Doyle: “Essentials of robust control”. Prentice Hall.
- [ZLC⁺18] J. Zhao, J. Li, Y. Cheng, L. Zhou, T. Sim, S. Yan, and J. Feng: “Understanding humans in crowded scenes: Deep nested adversarial learning and a new benchmark for multi-human parsing”. *Proceedings of the 26th ACM international conference on Multimedia*, 792–800, 2018.
- [ZWZ⁺21] Q. Zhu, Z. Wang, H. Zeng, T. Yang, and X. Wang: “Effects of graphene on various properties and applications of silicone rubber and silicone resin”. *Composites Part A: Applied Science and Manufacturing*, 142:106240, 2021.
- [ZXH19] Q. Zhang, Y. Xue, and X. Huang: “Online training for body part segmentation in infant movement videos”. *ISOEN 2019*, 489–492. IEEE, Piscataway, NJ, 2019.
- [ZZP⁺18] Z. Zhang, X. Zhang, C. Peng, X. Xue, and J. Sun: “Ex-fuse: Enhancing feature fusion for semantic segmentation”. V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss (eds.), *Computer vision - ECCV 2018, Lecture notes in computer science*, vol. 11214, 273–288. Springer, Cham, 2018.

Curriculum Vitae

12.11.2025	Doctoral Examination for Dr.-Ing.
2020-2024	Research Associate Medical Information Technology RWTH Aachen University, Germany
2019	Master of Science
2018	Exchange Semester Czech Technical University in Prague, Czech Republic
2016-2019	Studies of Electrical Engineering, Specification Biomedical Engineering RWTH Aachen University, Germany
2016	Bachelor of Science
2013-2016	Studies of Electrical Engineering, Specification Computer engineering RWTH Aachen University, Germany
2005-2013	General Higher Education Entrance Qualification (Abitur) Städtisches Gymnasium Haan, Haan, Germany
1996	born in Haan, Germany

