Welding in the World
https://doi.org/10.1007/s40194-024-01895-2

RESEARCH PAPER

Development and validation of a generalized, AI-based inline void defect detection solution for FSW based on force feedback

P. Rabe[1] · A. Schiebahn[1] · U. Reisgen[1]

Received: 7 August 2024 / Accepted: 19 November 2024

Abstract
Friction stir welding is a solid-state joining process that operates below the material's melting point commonly used to join aluminum parts, avoiding the drawbacks of fusion-based methods. These resulting advantages have accelerated growth and are increasing the number of applications across a range of industrial sectors, many of which are safety–critical. Along with the increase in applications and rise in productivity the need for reliable and cost-effective, non-destructive inline quality monitoring is rapidly growing. This publication is based on the research group's ongoing efforts to develop a capable generalized inline-monitoring solution. To detect and classify FSW defects, convolutional neural networks (CNNs) based on the DenseNet architecture are used to evaluate recorded process data. The CNNs are modified to include weld and workpiece-specific metadata in the classification. These networks are then trained to classify transient weld data over a wide range of welding parameters, three different Al alloys, and two sheet thicknesses. The hyperparameters are incrementally tuned to increase weld defect detection. The defect detection threshold is tuned to prevent false negative classifications by adjusting the cost function to fit the needs of a force-based detection system. Classification accuracies > 99% are achieved with multiple neural network configurations. System validation is provided utilizing a newly recorded weld dataset from a different welding machine with previously used parameter/workpiece combinations as well as parameter combinations and alloys as well as sheet thicknesses outside the training parameter range. The generalization capabilities are demonstrated by the detection of > 99.9% of weld defects in the validation data.

Keywords Friction stir welding · Quality monitoring · Deep neural network · DenseNet

1 Introduction

The presented research is embedded within the framework of the interdisciplinary, collaborative Cluster of Excellence "Internet of Production" of RWTH Aachen University. The research collaboration is aimed at the digitalization of production chains [1]. As a part of this extensive research project, this publication is a continuation of efforts in developing a generalized defect detection solution with a wide application range. The new developments are based upon previous insights, developments, and achieved results by the research group. The experimental methodology has been established before, allowing for direct comparison of results. The major new developments in this paper relate to the customization of the neural network used for FSW classification and the resulting increase in classification accuracy to 99.77%. Furthermore, the validation steps, including variations in methodology, welding machine, and welding task, are described. The results of the validation and achieved generalization are shown and related to industrial applications for defect detection. The fundamental motivation of the project and used methodology are constant along the development path and accompanying publications, which therefore overlap in the in the chapters regarding introduction, motivation, state of the art, and experimental setup [2–6]. Prior analysis approaches and results that build the base for the described developments can be found in previous publications by the same research group [2–6]. Since the last publication, the dataset has been increased to include further alloys and sheet thicknesses. The performed optimizations improved the defect detection

Recommended for publication by Commission III—Resistance Welding, Solid State Welding, and Allied Joining Process.

✉ P. Rabe
pascal.rabe@isf.rwth-aachen.de

[1] Welding and Joining Institute, RWTH Aachen University, Pontstraße 49, 52062 Aachen, Germany

Published online: 11 December 2024

accuracy by 1.40% to 99.77% through hyperparameter tuning. Furthermore, the detection threshold was tuned to better fit the application of process monitoring by replacing the cost function to successfully prevent false negative classifications. To provide system validation, the development was successfully tested on a completely new dataset from different machines, new alloys, and plate thicknesses, showing applicability by detecting > 99.9% of defects and challenges for further research [2, 6].

Friction stir welding (FSW) was patented by The Welding Institute (TWI) in 1991 [7]. It is a modern solid-state welding process that produces high-quality welds through intermixing in the plastic state and dynamic recrystallization, using frictional heat and pressure generated by a rotating, non-consumable tool. The challenges associated with conventional fusion welding processes of aluminum alloys are mainly avoided due to the solid-state nature of the process, and a fine microstructure is achieved. In conjunction with the reduced specific energy input, this yields superior mechanical and technological properties, as well as high process energy efficiency, making FSW a sought after joining process in the aerospace and rolling stock industries [8]. Furthermore, the growth of electromobility as well as increased efforts in lightweight vehicle construction has increased the number of FSW applications in the transportation industry. Many new developments and applications are centered around battery trays, heat exchangers, and mixed material joints of copper and aluminum for electrical power systems [9, 10]. In industrial production, quality assurance is provided through testing methods conducted after the joining process, which necessitate additional production steps and increase production time and cost. The increase and diversification of applications and the associated growth of production volume motivate the need for reliable and cost-efficient, non-destructive inline quality monitoring, which is easily applicable to the individualized production environment by the growing number of FSW applicants [11].

Specialized FSW welding machines capable of highly automated welding with closed loop axial force control to adaptively control the welding process are mostly used in current serial production. To utilize force control during the process, the machines are equipped with sensors to monitor axial force and often further welding parameters. However, between different manufacturers and machines, the sensors vary in their accuracy, recording frequency, and latency and do not provide sufficient process force feedback for high-quality monitoring of the processes' dynamic behavior. Due to the shortcomings of direct measurements and internal sensor systems, external sensors have been used to establish relationships between the weld process dynamic, recorded data, and weld seam quality, for example [2, 12–15]. These approaches generally utilize the analysis of cyclical occurrences of the dynamic force components of the in-plane welding force feedback or variation analysis of axial force and torque graphs. The cited sources generally achieve the goal of defect detection, despite varying sensors and analysis methods. However, the given examples are limited in their applicability as they are developed for a fixed joining task (single alloy, single sheet thickness, single welding tool) sometimes with a single parameter set.

Within this work, the generalization of quality monitoring between different FSW welding tasks is addressed by supplementing the recorded welding force and torque data with specific meta-data to enhance categorization accuracy. This demonstrates the generalization capabilities of densely connected neural networks across different alloys, sheet thicknesses, and a wide range of welding parameters. To further improve the applicability of the inline capable quality monitoring, the detection threshold is shifted to prevent false defect-free determinations by adjusting the network's cost function. The generalization of the quality monitoring capability is achieved by training deep learning networks based on DenseNet [16] architecture to analyze recorded weld data and locate subsurface volumetric defects. These voids are the most common defect in FSW and cannot be identified or located by visual inspection [17]. Therefore, detection requires destructive testing, or inspection by phased-array ultrasound or radiographic testing methods [8, 18]. The final development is then validated through further experiments, classifying weld data recorded on a different welding machine, using further alloys, sheet thicknesses, and welding parameters.

1.1 Material transport and defect formation during FSW

In industrial production environments, the FSW process is regarded as a comparably stable and well-controllable joining process, allowing for efficient implementation and reliable joining of components. However, the process is susceptible to a high number of external disturbances and variations in process conditions that can result in weld seam defects. Some of these defects are not detectable by the most widespread quality monitoring approaches, e.g., axial force monitoring and visual inspection [19].

The factors causing process disturbances can be divided into welding parameter–based factors, welding machine influences, and workpiece irregularities. When investigating weld quality, the machine-based influence factors can be regarded as constant and only need to be analyzed when developing generalized solutions, as these are fixed for the combination of machine, workpiece, fixture, and used welding tool [20]. Regarding influences, the welding tool is a special case, as tool wear needs to be monitored closely in long-term analysis and production environments, as wear significantly influences process forces and their dynamic behavior [21, 22]. In production environments, the main welding parameters, spindle speed, feedrate, and axial force

for each welding task are generally developed empirically and fixed for that task. As a result, significant disturbances are not caused by those, but by the interaction with the workpiece, which causes defect-inducing deviations. Gap tolerances, thickness variations, hardness gradients, and surface condition changes are the most common workpiece irregularities [23]. Changes in hardness, thickness variations, and gap tolerances influence the material volume underneath the welding tool and thereby the heat input and transport of plasticized material during weld seam formation, leading to process instabilities [23]. The surface condition of the workpiece and tool as well as tool wear have an impact on the interface contact condition, heat input, and material transport through wear, adhesion, and friction coefficient [24]. A deviation from the steady cyclical state can be caused by each factor or combination of these. The deviation can lead to a process state in which the employed quality control cannot detect defects occurring during the weld seam formation.

Related to the detailed influences, two main defect initiation mechanisms have been described. Plunge depth is a main factor in heat generation and process material transport and can be monitored through machine parameters and related to weld quality. Change can be caused by a variety of factors, leading to either increased or reduced tool plunge depth, resulting in incomplete penetration and decreased weld properties [25]. Increased depth can result in close proximity or contact of tool and backing plate, causing adhesion between workpiece and backing, tool failure, or further defects induced by excessive temperature and abnormal material transport.

The second identified defect initiation mechanism is significantly more difficult to detect by current quality monitoring solutions, as it is caused by irregular material flow in the stir zone. The periodically oscillating process condition and resulting intermittent material transport around the tool can be influenced by changes in the specific energy input due to changes in surface condition, workpiece strength/thickness, interface condition, or tool wear, thereby disrupting the weld seam formation [26, 27]. Insufficient material transport can cause local or prolonged internal volumetric defects in FSW weld seams, such as voids, cavities, or surface defects. Due to these occurrences, FSW production using current process control and quality monitoring requires further quality inspection of the FSW weld seams. The additional quality control measures add complexity, time, and cost to the production.

1.2 State of FSW process monitoring developments and weld seam quality assessment

Heating through friction at the tool-workpiece interface and translational movement of the tool through the plasticized material result in comparatively high process forces, which are often used to characterize the FSW process. During the steady welding state, process forces consist of static and dynamic components present in all three spatial directions in an oscillating cyclical manner corresponding to the rotational speed of the welding tool and its higher harmonics [28]. The cyclical oscillation of the spatial forces at the frequency imprinted by the spindle is widely accepted; however, many further influence factors, interdependencies, and relations to force magnitude, the occurrence of higher harmonics, material transport, and dependence on welding parameters, tools, and workpiece properties are not fully understood or quantified. Many publications have been centered around establishing correlations between these oscillations, their deviation, and FSW weld seam quality, despite their unclear origin. Multiple different approaches to relate weld quality to force deviations and torque oscillation have been published, e.g., [3, 6, 12, 13, 29–31]. The cited works are based on the proven assumption that all major process influences can be summarized in the recorded process force feedback and thereby a representation of the process dynamic behavior and introduced deviations and instabilities can be drawn. From this compounded representation of the process state, an indication towards weld quality can be derived.

Multiple publications show the possibility of empirical anomaly detection based on welding force evaluation. These evaluations are mostly based on extracted incremental features or gradual force changes within the recorded process data. The majority utilizes changes in the dynamic components or combinations of changes in the static and dynamic force compounds. The successful correlation of weld defects by Jene is however based on the mean average lateral force [29]. Further research into the empirical evaluation of process forces [12, 31] proves the feasibility of the evaluation of dynamic components, but also portrays the limitations and the need to adjust the evaluation and selected features for every change in welding conditions, limiting applicability. Analytical correlations can also be formed in the frequency domain, evaluating frequency spectra. Internal void defects can be indicated by high amounts of low-frequency oscillation, according to Gebhard [20].

Frequency domain data was also used by Boldsaikhan et al. in the first application of artificial neural networks (ANNs) to detect internal welding defects in FSW [32]. Torque data recorded at 51.2 Hz was transformed into the frequency domain and served as the feature vectors to train fully connected neural networks (FCNNs). The dataset used was heavily biased towards defect-free welds, and the FCNN was able to correctly identify internal weld defects in the test dataset. Based upon the results, subsequent work by the main authors shows the use of wavelet transformations of the in-plane forces to evaluate weld quality with

ANNs [33]. The trained FCNN reached an accuracy on the test dataset of 95% for internal defects > 0.08 mm. Further developments of AI technology and especially deep learning techniques allowed Hartl et al. [34] to classify welds through weld surface images and weld force data using convolutional neural networks (CNNs). The highest classification accuracy of 79.2% was achieved with a CNN based on Alex-Net [35], when evaluating the lateral force (F_y).

While different ways to correlate the dynamic process behavior to weld quality have been proven successful, none of the published works thus far can generalize the quality prediction across different workpiece materials, thicknesses, and multiple welding tools. Most investigations limit the ranges of feedrate and spindle speed to values oftentimes not applicable to efficient industrial production. However, the possibility of detecting internal weld defects through force feedback analysis was proven. The published progress thus far does not provide an applicable solution for industrial production requirements. The necessary training datasets for each welding task require extensive effort to produce, prepare, and label through a large number of test welds with sufficient amounts of variation, due to the inability to generalize across different welding tasks.

In this study, the application range and applicability of inline defect detection is increased by using deep learning densely connected CNNs (DenseNet) to examine the transient temporal sequences of FSW process forces and axial torque. The force recordings are supplemented with weld meta-data to promote generalization and feature scaling and thereby enable the classification of welds with internal defects. Furthermore, the CNN cost function is adjusted to prioritize false defect detection to prevent non-detection of defects. A dataset of classified weld segments of full penetration welds over a wide range of feedrates and spindle speeds, five Al alloys, and four sheet thicknesses is established for the neural network training and testing. Micro-focus radiographic testing (μ-RT) pictures are analyzed to classify the welds and label the weld data. Subsequently, DenseNet-CNNs are trained to extract features and classify the welds based on force measurements, torque recordings, and the supplemented weld meta-data. The generalization and limited extrapolation capabilities of the developed solution are then successfully validated using weld data recorded on a different FSW machine with an extended range of alloys, sheet thicknesses, and welding parameters.

2 Experimental setup and data acquisition

The welding experiments were performed on two different welding machines. Welds for the training dataset were executed on a moving table portal type FSW machine Type 345C, built by Precision Technology Group Ltd (PTG). Due to its design and fabrication, it provides position-independent low static compliance. The vibration excitation response is comparably low, minimizing the influence on the process dynamic behavior [36]. Weld data for the validation dataset was recorded of welds performed on a console stand design FSW machine by Aerospace Engineering Equipment (AEE), type HT-JC06 × 08. The build type has position-dependent compliance and overall lower stiffness [36]. On both welding machines, the workpieces were clamped using the same fixture. The backing plate was made from 8-mm mild steel that had been ground and subsequently artificially oxidized to prevent workpiece adhesion. The workpieces are 135 × 300 mm^2, allowing for 260 mm weld length. For the training dataset, workpieces of 1.5 mm AA5754-H22, 3 mm AA5754-H22, 1.5 mm AA7075-T6, and 1.5 mm AA6082-T6 were used. For the validation dataset, workpieces made from 1.5 mm AA5754-H22, 1.5 mm AA7075-T6, and additionally 1 mm AA2017-T651 and 2 mm AA5083-H111 were used. To prevent the influence of edge shape and gaps between plates, the welds were performed as full penetration seam on plate welds. The geometric tool parameters were adjusted to the workpiece thickness. The tools were fabricated from H13 tool steel as monolithic components and hardened for the experiments. The tool geometry and relevant dimensions are shown in Fig. 1 and Table 1. A tool tilt angle of 1.5° was used during all welds.

The investigated parameter ranges of welding feedrate and spindle speed (RPM) were set to evaluate a wide range of industrially applicable parameters within the bounds of the measurement device range and machine capabilities. The weld parameter combinations were chosen to investigate the driving variables of measured force feedback and resulting deviations in welds containing defects for the range of combinations of sheet thicknesses and alloys. The main force excitation at the frequency of the tool rotation speed is mainly driven by the superposition of radial tool runout and discontinues material transport, as well as

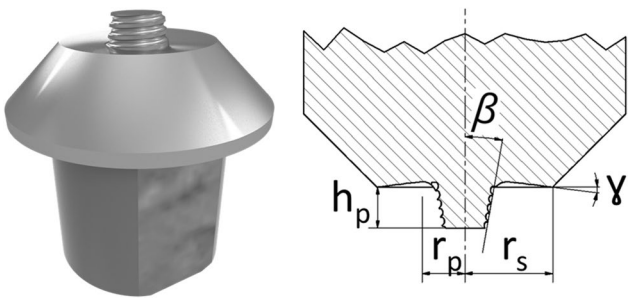

Fig. 1 Rendering and cross-section drawing with relevant dimensions of used tools (cf. Table 1)

Table 1 Relevant dimensions of welding tools

Parameter	Tool 1.0 mm	Tool 1.5 mm	Tool 2 mm	Tool 3 mm
Probe radius, r_p	2 mm	2 mm	2.5 mm	2.5 mm
Shoulder radius, r_s	6.0 mm	6.0 mm	6.0 mm	7.0 mm
Conical probe angle, b	10°	10°	10°	10°
Probe length. h_p	0.92 mm	1.42 mm	1.92 mm	2.85 mm
Concave shoulder angle, g	7°	7°	7°	7°
Pin feature	Metric thread	Metric thread	Metric thread	Metric thread

potential tool pin features [37]. The amplitude of the oscillation is driven by the temperature-dependent material resistance to the extrusion of the welding zone by the tool, as well as the tool path and parameter-dependent extrusion volume. The welding parameter combinations of the training dataset were chosen to investigate these influences. A total of 13 parameter combinations was set in three groups of five combinations each with different governing variables, intersecting another group in one combination. For the validation experiments, four parameter combinations from the training data were chosen and two combinations, one within the training parameter window and one outside of it, were added. The parameter combinations in their distinctive sets are shown in Fig. 2. The first set has a fixed pitch, the relation between feedrate and tool revolution, to allow the analysis of increasing feed and axial force due to increased feedrate and thereby reduced thermal softening in front of the weld. Sets 2 and 3 are centered around fixed spindle speeds with increasing feedrates to retain the main excitation frequency across the set with increasing amplitude and deviations.

A total of 1200 RPM is the chosen spindle speed for set 2. It offers a high number of measurements per tool revolution (125 as detailed below) and allows for a wide range of feedrates without defects. This enables the accurate measurement of welding forces over a wide range of parameter combinations. The spindle speed of 1800 RPM was chosen for set 3 according to the dynamic frequency response of the used PTG FSW machine [36]. The first higher harmonic of the 1800 RPM is 60 Hz, which is equal to the lowest identified natural frequency and should thereby increase force oscillation in lateral direction and detectable variations within the force signal, despite the pre-tensioning through the axial welding force. The range of the feedrate was determined by the capability of producing defect-free welds. The lower limit was set to 600 mm/min so that no overheating or surface galling occurred. The upper limit was set by the feed-force limit of the measuring device used to record the weld data. For set 1, the minimum feedrate could be lowered due to lower spindle speed, and the maximum feedrate was limited by the machine's maximum spindle speed of 3000 RPM.

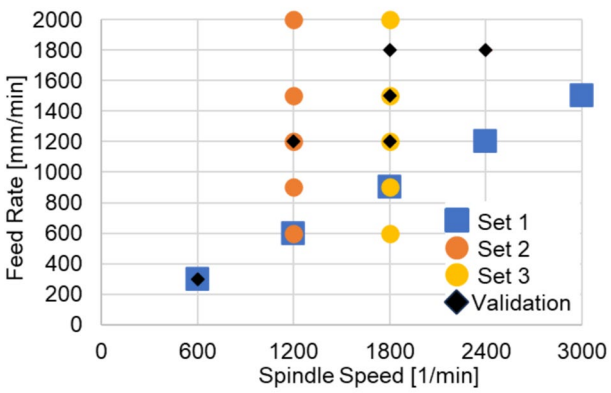

Fig. 2 Welding parameter combinations of sets 1–3 and validation experiments

FSW is generally performed with closed-loop force control in industrial production environments to enable the compensation of workpieces and fixture variance. For these welds, position control mode was chosen, as it allows for reliable and reproducible plunge depth relative to the machine coordinates, eliminating the influence of depth variations from the recorded data. It furthermore prevents the influence of lag, set force deviation, machine control imprinted z-axis oscillation, and drift on the recorded data by eliminating the closed loop force control system. This decrease in systemic variance enables reliable and reproducible production of defect-free welds and welds with inner volumetric defects for each parameter combination. To achieve defect-free and defective welds, two plunge depths were determined for each parameter combination and alloy during pre-trials. The reduced plunge depth was set to decrease heat input and forging pressure to produce welds with inner volumetric defects, but without any detectable surface defects. Each parameter combination of the training set was repeated three times with each plunge depth to provide a sufficiently large dataset with statistical validation. For the welds of the validation set, two plunge depths were determined, one to result in defect-free welds and the other to produce inner defects. For the validation experiments, the plunge depths were not used for individual welds,

but rather combined through a plunge depth ramp within each weld. For each combination, both ramp directions were repeated twice. Figure 3 shows the plunge depth ramps and the resulting axial forces for a weld performed at 1200 mm/min in 1.5 mm AA7075. This methodology was chosen for the system validation dataset, as it resembles the defect initiation of production environments more closely than the previous method which produced mainly uniform defective or defect-free welds.

The force and axial torque data is recorded during the welding process for analysis by a sensor unit integrated into the tool holder, Spike® mobile. The positioning of the measurement system within the tool holder makes the recording independent from the workpiece fixture; however, the forces in all spatial directions, F_x, F_y (in welding plane), and F_z, are recorded relative to the welding tool and not the machine axis. F_x and F_y are calculated from the bending moment and lever length between the tool shoulder and measuring point. Torque and axial force are measured directly by strain gauges placed at a tapered part of the tool holder. The signal of each data channel is logged at 2.5 kHz and wirelessly transmitted to a receiver integrated into the machine control system and the process control. The data is then transmitted within a wired network containing edge devices and servers for storage, processing, evaluation, classification, and decision support. The signal measurement and data processing chain are shown in Fig. 4. The tool position in the machine coordinate system and process data is logged and stored at 500 Hz, besides the process force and torque measurements. Furthermore, at the start of each weld, the welding program and its parameters are recorded to amend weld meta-data to the process recordings.

3 Results and discussion

3.1 Welding results, classification, and datasets

The welds were performed on two different welding machines as described above and shown in Fig. 2 and Fig. 3. Some weld parameter and alloy combinations could not be welded without surface defects or exceeding the force limit of the measuring device; therefore, the total number of welds and corresponding measurement data for the investigations was reduced to 281 welds for the training dataset and 99 welds for the validation dataset. After welding, all welds accepted to the datasets passed visual inspection without identification of surface flaws. The welds were analyzed by radiographic testing (rt) to quantify the welding results and determine and localize internal weld seam defects. Radiographic testing was performed on a micro-focus computed tomography machine (µ-ct), Type Viscom XT9225. In micro-focus tomography, the X-rays originate in a single spot source on the target with opening ray paths towards the detector. The magnification of the specimen is achieved by placing it in the opening ray path between the source and detector, and the rate of magnification is determined by the chosen distances and is directly proportional to the detectable defect size [38, 39]. In order to acquire high-resolution images with a low defect detection threshold on the comparably long and narrow welds, two pictures were taken of each weld, one for the first half and one for the second. The digital images were then compounded using the added wire indicator as a reference. A duplex wire indicator was used in accordance with ISO 19232–5 to accurately determine the detectable size of volumetric defects [40]. The detection threshold for voids and tunnel defects (internal volumetric

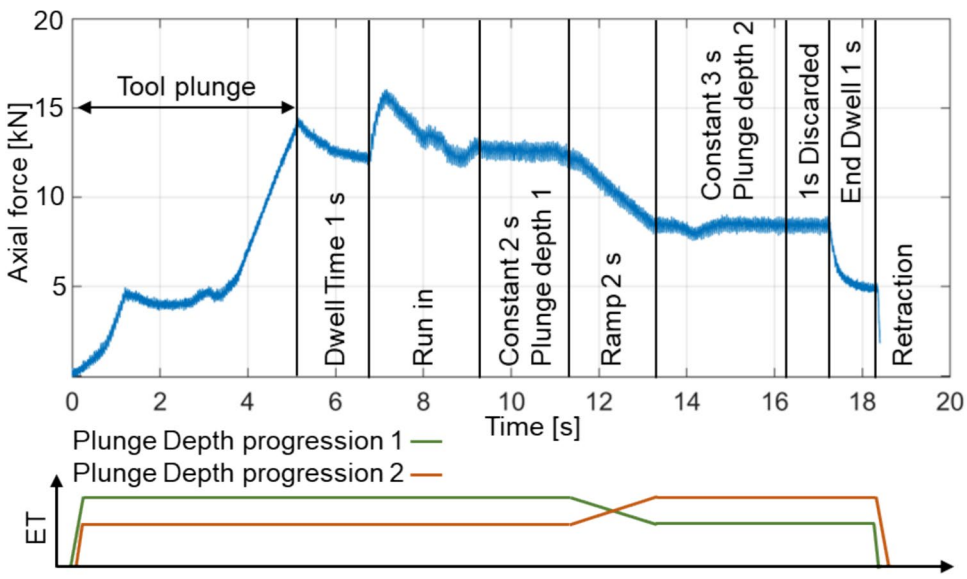

Fig. 3 Axial-force graph with plunge depth ramps within the measuring duration for a weld performed at 1200 mm/min

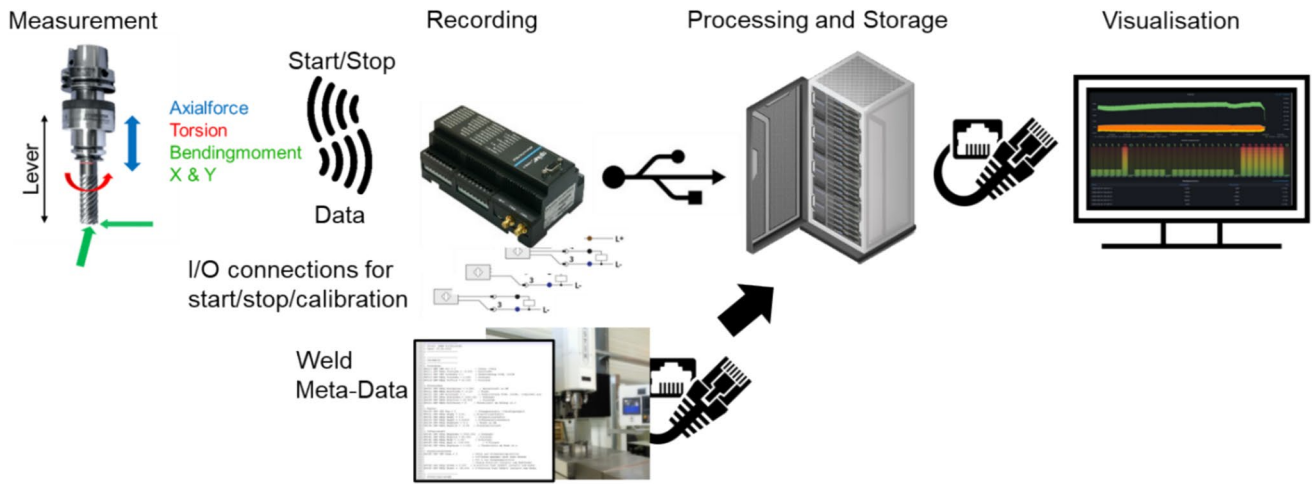

Fig. 4 Weld data measurement, processing, recording, and visualization

defects) was determined to be 0.063 mm (D12) for 1-mm and 1.5-mm thickness plates and 0.08 mm (D11) for 2-mm and 3-mm plates. Thereby, the set detection thresholds exceed those commonly achieved by quality control measures in industrial applications, excluding microsections, which only allow for local determination. To achieve the low detection threshold, the brightness and contrast of the pictures were adjusted in image post-processing. For weld classification, the adjusted images were manually analyzed to detect, locate, and mark internal weld defects. During analysis, the plunge and tool exit locations were excluded. The analysis showed 127 welds (45,20%) with defects and 154 welds (54,80%) without defects for the training dataset and 93 of 99 (93,94%) welds containing defects in the validation dataset, thereby showing slightly less defects than targeted in the training data welds and some defect-free welds despite the plunge depth ramp in the validation experiments. The rt-image analysis was validated through the analysis of micro-sections. Multiple cross-sections were taken from different locations including defective and defect-free areas. The validation was in perfect accordance with the rt-image analysis and is therefore used for the weld categorization of both datasets.

For the data analysis, the development of an inline quality monitoring solution, and the capability of localizing defects through data analysis, the welds and their recorded process data are split into shorter sections. For each second of recorded data, a new section starts, lasting for 3 s of data. The sections thereby overlap the previous and subsequent sections by 1 s each. The segmentation is shown in Fig. 5 for a weld performed at the maximum welding speed of 2000 mm/min. The overlapping segmentation increases training data and allows the same data to be evaluated in multiple data contexts, generating better generalization in feature and threshold determination.

The 3-s long sections of data result in 7500 measurements for each spatial force direction and tool torque at the measurement frequency of 2.5 kHz, giving 30,000 data points for analysis per segment. The weld length and highest feedrate were used as references to determine the maximum number of sections for each weld, resulting in five sections per weld. At lower feedrates, the sections were taken from the end of the weld sparing the last second of the weld as well as the tool exit location, as shown in Fig. 3.

According to the rt-images and the defect identification and localization, the individual sections of the welds of the training and validation set were classified. Classes were determined based on the existence of internal void defects in the section (NOK) and sections without any detectable defects (OK). The classification according to the identified threshold is shown in for the training dataset and in for the validation dataset. For the NN training and evaluation, the class determination was transferred to a weld segment classification in which class 1 corresponds to NOK welds, signifying the detection of a defect, and class 0 corresponds to OK welds, determining that no defect exists within the weld seam segment. The final dataset sizes and determinations are given in Table 2 for the training dataset and Table 3 for the validation dataset.

3.2 Selection, modelling, training, and testing of neural networks

Through an increase in understanding of machine learning approaches and development for numerous applications, many new and efficient deep learning neural network architectures have been developed. The economical access to high-performance computing and network architectures adapted to various tasks increases the benefits of recording production process data and eases the advantageous analysis

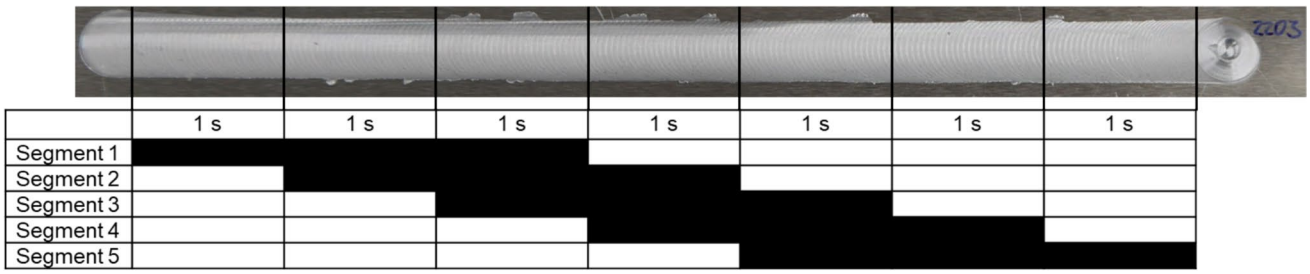

Fig. 5 Segmentation of weld data at maximum welding speed (2000 mm/min)

for manufacturers [41]. Based on previous developments, densely connected CNN (DenseNet) has been chosen for this work [6]. The dense connections of the teachable layers allow for the detection, extraction, and classification of low-, mid-, and high-level features during network training and inferencing [16], matching the dynamic behavior and defect-induced variations of FSW weld data. Deep learning eliminates the brittle, time-consuming, and not scalable hand-engineered feature extraction of empirical analysis and other AI methods based on manual feature input. Empirical solutions for FSW quality monitoring have been disproven by using different datasets. The versatile architecture of DenseNet combined with automated feature detection allows for the classification of diverse input data [42]. DenseNets detect all complexity levels of features by propagating each convolutional layer output feature map to all consecutive layers. This "collective knowledge" from all preceding layers enables the classifier to evaluate all complexity-level features and enhances feature reuse. The propagation and stacking of feature maps as input for the next convolution are shown in Fig. 6. The advantages of using all feature complexity levels for data evaluation have been proven in computer vision problems and transfers to FSW data, as it contains multiple channels of oscillating data with multiple excitation frequencies and irregularities related to weld defects.

The recorded process forces in all three spatial directions and the tool torque were used as the main inputs for training and evaluation of the generated and labeled dataset, allowing for supervised learning considering the weld classification illustrated above. Weld force response data was chosen for evaluation, as it summarizes all influence factors and accurately depicts the FSW process dynamics and resulting weld seam quality in a limited number of data streams. The generated dataset, described as training data, was used to train the deep neural networks to generate feature maps and detect characteristic features to classify the welding force response data to identify internal weld seam defects. Deep learning was chosen to automatically detect features that exceed manually identified and engineered features. Deep learning replaces the manual feature detection through multiple layers that weigh the inputs, sum the

Table 2 Classification of weld seam segments of the training dataset

Training data	Full set	AA5754 1.5 mm	AA7075 1.5 mm	AA6082 1.5 mm	AA5754 3 mm
OK: defect-free segments	788 56.09%	243 60.75%	211 51.46%	233 59.74%	101 49.27%
NOK: volumetric defect segment	617 43.91%	157 36.25%	199 48.54%	157 40.26%	104 50.73%
Sum	1405	400	410	390	205

Table 3 Classification of weld seam segments of the validation dataset

Validation data	Full set	AA5754 1.5 mm	AA7075 1.5 mm	AA5083 2 mm	AA2017 1 mm
OK: defect-free segments	184 37.17%	78 65.00%	34 28.33%	8 6.40%	64 49.23%
NOK: volumetric defect segment	311 62.82%	42 35.00%	86 71.67%	117 93.60%	66 50.77%
Sum	495	120	120	125	130

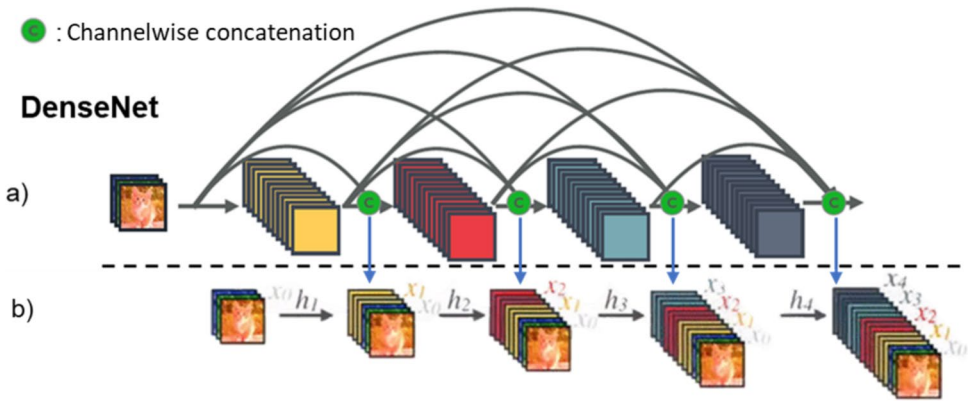

Fig. 6 Representation of dense connectivity. **a** Propagation of inputs and feature maps with channel-wise concatenation. **b** Concatenation of feature maps to all previous as input to the next layer

weighted input, and apply non-linear activation functions to generate an output [42]. The shared weight architecture of the convolutional filters makes the detection of classification relevant features space and orientation invariant allowing for recognition and categorization across the evaluated data. For this analysis, a densely connected neural network with 716 layers was chosen, based on DenseNet201 [16]. The dense connections enable the classification of all feature complexities and furthermore aid in error backpropagation, allowing for faster convergence. The network consists of several "DenseBlocks," which are densely connected and control the growth rate of the network. The blocks are made up of 6–48 repetitions of the sub-block centered around the convolution. The blocks are connected with transition layers in between, which reduce the width of the network. All convolutional layers use pre-activation batch norm to apply normalization to enhance performance. The base architecture of DenseNet201 was modified to increase the generalization capabilities and allow the CNN to extrapolate beyond its training parameters. In order to expand on the capabilities, additional feature inputs were used. The feature inputs used meta-data of the welding task, including the welding parameters (spindle speed, feedrate) and workpiece parameters (sheet thickness, flow stress at 450 °C) to accurately scale the detected features according to the current state of the evaluated weld data. The welding parameter input was chosen, as it represents the two most important set welding parameters for the weld and drives the oscillation frequency and force level. The sheet thickness input allows the features to be scaled according to the required process forces. The flow stress at 450 °C scales the forces required for material transport at or around the working temperature of the process. It is also widely available for Al alloys as it is required knowledge for extrusion production processes. The amended form of the used CNN is shown in Fig. 7.

For the DenseNet training, the recorded data (F_x, F_y, F_z, Torque) of each segment was used as the input feature vector, along with the weld meta-data. Different configurations for the data input were investigated, as well as a variety of

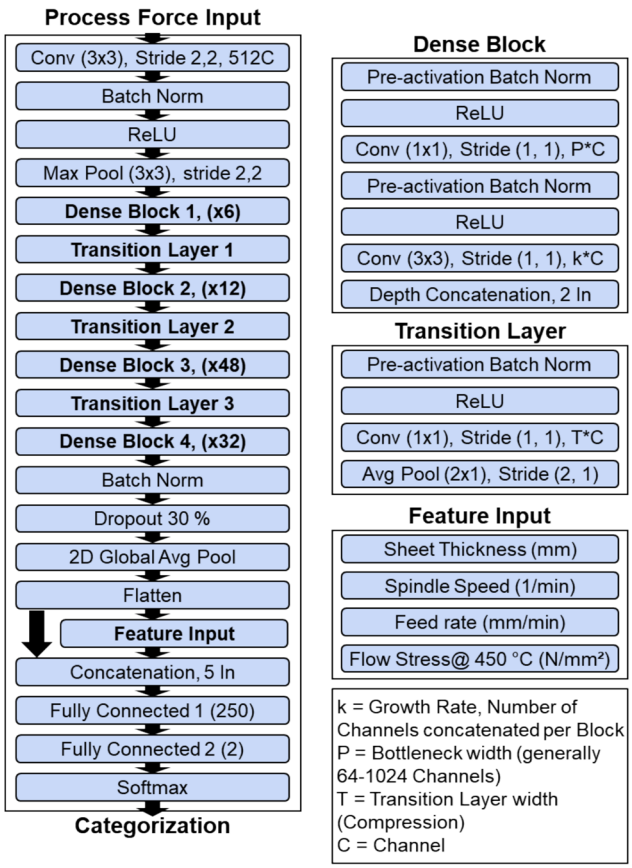

Fig. 7 Network architecture of the modified DenseNet-CNN

growth rates, bottleneck width, and transition layers. For these, the hyperparameters were incrementally tuned resulting in > 100 configurations. The investigated parameter ranges are given in Table 4.

Besides the network configuration and the hyperparameters, the cost function, calculating error values and therefore network updates, was also adjusted to match the objective of the application. Generally, in classification tasks, the network is optimized to minimize the number of resulting

classification errors, by applying equal penalties to false positive and false negative classifications and thereby moving the network towards optimization. The system developed within this paper aims to serve as quality control for FSW; therefore, the even distribution between false positive and false negative errors reduces applicability, by passing defective welds as non-defective. In order to prevent the false classification of defective welds as defect-free, the cost function is adjusted to strongly penalize this classification category and thereby adjust the generalized defect detection threshold of the network. The original and adjusted cost functions for the network outputs y and targets t are given below.

$$\text{Original cost function: } d(t,y) = -\frac{t}{y} + \frac{(1-t)}{(1-y)} \tag{1}$$

$$\text{New cost function: } d(t,y) = -(\mathrm{t}-\mathrm{y}) * \left(\mathrm{y} * a_1\right) - (\mathrm{t}-\mathrm{y}) * \left(\mathrm{t} * b_1\right) + 1 - \mathrm{t} - \mathrm{y} \tag{2}$$
$$a_1 \lll b_1$$

The given formula shows the addition of two new hyperparameters, which optimize the proportion of network penalty and error value between the categories of false classification and thereby the defect classification threshold. The results of the change of the cost function and the influence of the new hyperparameters on the calculated error values are shown in Table 5. To determine the success of the threshold modification for process quality evaluation, the positive factor (P_f) is introduced, and it relates the number of false negative classifications to the number of false positive classifications. It is established this way, as division by the targeted number of 0 false negatives is not possible.

$$\text{Positive factor: } P_f = \frac{\text{number of false negative}}{\text{number of false positive}} = \frac{t-y=1}{t-y=-1} \tag{3}$$

Different data input configurations had a substantial influence on the training duration and accuracy. It was determined that 4 separate nets, 1 for each data stream, concatenated before the fully connected layer, overfit to the data for all hyperparameter configurations. A 2-D representation of the data as a 7500×4 matrix, whereas the data was normalized by channel prior to the merge, vastly increases computational time due to the second iteration dimension for the applied filters. The best input configuration was determined to be $7500 \times 1 \times 4$, giving 4 separate channels of 7500 by 1 data into the network per segment. For this configuration, different network growth rates (k) and transition layer width (T) were tested and the hyperparameters were incrementally tuned. Their interdependencies were monitored during tuning, and overfitting was reduced by tuning momentum and L_2 regularization. A major improvement was achieved by tailoring learn rate and learn rate decay to the training dataset and number of training iterations.

Network training was performed on a consumer workstation PC with a GPU for training acceleration. On a last-generation consumer GPU (Nvidia RTX 3090 with 24 GB VRam), each training iteration for the full dataset took 0.75–1.5 h, depending on network setup. Each network configuration and hyperparameter combination was repeated five times, and the top 3 results were used for evaluation. This was done to prevent bias due to unbalanced datasets and unfavorable initialization. For each training iteration, the training dataset was split randomly into 80% training data, 10% validation data, and 10% testing data. The training data is used by the network to detect classifying features by calculating and updating weights and biases. The test set data is exempt from the network training and afterwards used to test the classification accuracy of the fully trained network. The resulting classification accuracies can vary with the random allocation of data to the sets and the weight initialization. During training, four different network configurations were found to consistently achieve the highest classification accuracies. These were built around the two lowest network growth rates and different transition-layer widths. The network configurations and the average resulting test set classification accuracy and positive factor for $n = 3$ iterations are given in Table 6, relating to Fig. 7.

The achieved accuracies show the best results for the network configuration with the lowest number of learnable parameters (growth rate, transition-layer width), at 99.77% ($n = 6$) classifying 844 of 846 test set segments correctly, without a single false negative classification ($P_f = 0$). The wider networks showing slightly lower accuracies are believed to result from slight overfitting and the spread of features across an increased number of feature maps.

Overall, the six iterations of each of the four network configurations (24) misclassify 21 segments of weld data, with 20 false positive and 1 false negative classification. This proves the success of the cost function adjustment. A closer analysis of the segments shows that during training, only 18 different segments were misclassified, with 3 segments being misclassified by two separate training iterations. The incorrect classifications are also seam-dependent, with two consecutive segments of the same seam from three different seams being incorrectly classified as containing a defect. The challenge of permanent correct classification with a precise definition of the defect detection threshold for a wide range of applications can also be seen from the misclassifications. Despite the low variance within the weld seams, 4 of the 21 misclassified segments (19.05%) are at the transition

Table 4 Ranges of hyperparameters and no. of filters during network optimization

Parameter	Value range
Data input configuration	7500 × 1 × 4, 7500 × 4 × 1, 7500 × 1 × 1 in 4 separate Nets
Input data scaling	[0, 1], none
No. epochs	15–40
Mini batch size	8–60
Learn rate	0.0000005–0.01
Learn rate reduction	0–0.75
Learn rate reduction period	2, 3, 4, 5, 10, 14
Momentum	0.1–0.85
L2 regularization	1×10^{-6}–0.9
Network growth rate (k)	16–64
Size initial convolution	3 × 3, 4 × 4, 7 × 7
No. intial filters	64–4096
No. channels bottleneck (P)	64–2048
No. channels transition layer (T)	T1, 64–512; T2, 128–1024; T3, 448–3584
No. neurons in FC1	100–1000
Cost function	Cross entropie, custom

Table 5 Results of the cost function

Cost function (t, y)	Cost (old): target, discrete value	Cost (new): discrete value
Cost (0, 0)	1, 1	1
Cost (1, 1)	−1, −1	−1
Cost (0, 1)	∞,4.5036×10^{15}	a_1
Cost (1, 0)	$-\infty, -4.5036 \times 10^{15}$	$-b_1$

point between defect and no defect, i.e., to one side of the relevant segment is a segment with a defect, and to the other side is a segment without defects. Further analysis of the graphed data gives no indication for the misclassifications. Analysis of the welding parameters, alloys, and sheet thicknesses does not provide helpful insight for improvement, as it matches the distribution within the training dataset. The only exemption is the "coldest" weld parameter combination (1200 RPM, 2000 mm/min), from which segments of three alloys were misclassified. This highlights the limitation of the NN at the extreme limits of the training parameter window and emphasizes the challenge of extrapolation to higher forces and force variations outside the field of training data, without additional scaling of features.

3.3 Validation through extended welding experiments

To validate the intended generalization ability of the developed weld quality monitoring solution and prove the enhanced extrapolation ability, contrary to inherent NN limitations, enabled by the meta-data feature input, the above-described validation dataset of 99 welds with new alloys, sheet thicknesses, and welding parameter sets, welded on a different welding machine with new tools, is used. The welds are segmented like before, discarding the last second of weld data and utilizing 5 segments per weld seam for the dataset. Inferencing is performed with all trained network configurations and iterations shown in Table 6. The average results of the six iterations of each configuration are shown in the final row and column of Table 7. Within the table, the results are separated by alloy/sheet thickness and position of the weld parameter combination relative to the combinations of the training dataset (see Fig. 2).

The classification accuracy drops significantly (6.28–8.59%) compared to the network training, when inferencing the validation set data with the DenseNets successfully trained on the training dataset. The highest classification accuracy is achieved by the K16T + configuration at 93.13%. Individual iterations of the same network configuration show high reproducibility and low standard deviation across all network configurations. Contrary to the initial network training, a higher filter count and number of feature maps result in better classification accuracies for both investigated network growth rates. Both wider networks show more reliable, better generalization. This is based on the higher number of detectable features for classification and the wider spread of individual features over multiple feature maps due to the increased linearization in the wider networks.

The differentiation of the inferencing by parameter class shows some unexpected results. Especially, the high classification accuracy of AA5083, which was not part of the

Table 6 Best network configurations and resulting average classification accuracy ($n=3$)

Network configuration	K16T32	K16T +	K32	K32T +
Initial convolution	512 3×3	512 3×3	512 3×3	512 3×3
Transition layer T1, T2, T3	128, 256, 896	512, 512, 1024	128, 256, 896	512, 512, 1792
Bottleneck width (P)	128	128	128	128
Training duration (min)	41	49	57	80
Positive factor (P_f)	0/2	0/4	0.5	0/2
Classification accuracy (%)	99.53	99..05	99.29	99.53
Repetition of training				
Positive factor (P_f)	0/0	0/1	0/4	0/5
Classification accuracy (%)	100.00	99.76	99.05	98.82

training dataset, is surprising. However, it proves the influence of feature scaling, as sheet thickness and high-temperature flow stress are within the bounds of the training dataset. For AA2017, with flow stress within the bounds of the training parameters, but sheet thickness outside of the training data range, the accuracy of parameter class 1 is comparably low. This is assumed to be related to the lower process forces of the thinner material, outside the training force range. The conclusion is supported by the classification of classes 2 and 3 that is ≥ 94.5%, as the feedrate of classes 2 and 3 is 1800 mm/min and process forces therefore are higher. For the alloy used within the training dataset, a comparably low accuracy is received. However, the parameter-based generalization of classes 2 and 3 works well, resulting in accuracies comparable to class 1.

Based on the significant number of misclassifications in the validation inferencing, the misclassified segments are analyzed closer. Depending on the alloy, the source for a large group of misclassifications can be identified. Especially for the alloys used in the test set, most misclassified welds are positioned at the defect initiation site or the end of a defect. The sum of all segments misclassified by all iterations of the same network configuration ($n=6$) is given in Table 8, separated by network configuration and alloy. Furthermore, the sum and fraction of segments of the defect transition area, where the segment borders on segment including defects and another without defects, are given.

A large proportion (avg. 79.71%) of the comparatively high classification inaccuracy of the validation dataset can be explained by the analysis above. The fraction of misclassified segments that lie within the plunge depth ramp and are surrounded by a segment with a defect and a segment without a defect is high. This means that the area of defect initiation or defect end lies within the misclassified segment, and it includes sections of both classifications. Thus, a minimal shift of the features used for defect detection causes the misclassification of the segment. This is clearly recognizable, as the described network classification behavior occurs predominantly in the alloys used during network training. For these, 77.78–92.11% of the misclassified segments are assigned to this transition area. For AA2017, this is only 54.44–74.14%.

Furthermore, a second major influence on the classification accuracy can be determined when evaluating AA5754 misclassified segments. For the validation experiments, a new tool, with the same major geometry, but individual production tolerances, was used and showed significant influence on the process force dynamics, especially on the slowest feedrate welds (300 mm/min, 600 RPM). This can be identified when comparing the data of defective and defect-free segments of the same parameter combination from the training and validation set, and it leads to the incorrect classifications of a high number of AA5754 segments for the given parameter set. The slow welds account for up to 48.19% (K32) of the misclassified AA5754 segments (K16T32, 27.66%; K16T + , 36.1%; K32T + , 21.53%). This emphasizes the influence of the welding tool geometry and tolerances as well as the welding machines' static and dynamic behavior. It also offers an area of improvement for the generalization, where the developed system can be further enhanced by implementing additional feature inputs relating to the welding machine and distinct tool geometry.

Both identified major drivers for misclassification show the complexity of determining a generalized threshold for defect detection. It is shown that classification close to the threshold is a major challenge, that, given the welding state, can be detrimentally influenced by tool production tolerance. A further increase in dataset size should stabilize the determined detection threshold and refine classification features. Furthermore, based on the observations regarding network configuration, wider networks with features spread over a multitude of feature maps should increase generalization and support extrapolation through features scaled by additional meta-data inputs.

Table 7 Average classification accuracy of the validation dataset, subdivided by network configuration, workpiece material, and weld parameter combination

$N=6$	Param	No. of segments per iteration	K16T32	K16T+	K32	K32T+	Avg. Acc
Acc. AA2017 1.5 mm (%)	1	85	87.60	88.82	85.49	91.57	88.38
	2	25	95.33	96.00	93.33	93.33	94.50
	3	20	90.83	95.00	95.00	95.83	95.87
Acc. AA7075 1.5 mm (%)	1	80	88.33	93.13	88.13	89.58	89.79
	2	25	91.33	91.33	86.67	91.33	90.17
	3	20	89.17	87.50	83.33	84.17	86.04
Acc. AA5083 2 mm (%)	1	80	99.79	99.58	98.33	98.75	99.11
	2	25	100.00	100.00	100.00	100.00	100.00
	3	20	96.67	100.00	95.83	99.17	97.72
Acc. AA5754 1.5 mm (%)	1	80	85.21	87.29	86.04	87.71	86.56
	2	20	89.17	98.33	92.50	94.17	93.54
	3	20	91.67	92.50	87.50	91.67	90.84
Avg. classification accuracy			91.18	93.13	90.20	92.50	91.75

Parameter class 1: Welding parameters equal to training-data welding parameters

Parameter class 2: Parameter combination within the field of training-data parameter combinations

Parameter class 3: Parameter combination outside the field of training-data parameter combinations

4 Application for inline quality control

The resulting network classification accuracies of the training dataset vastly eclipse industrially applied defect detection methods such as ultrasonic testing, radiographic testing, and visual inspection, in not only detection accuracy and size of detectable defects but also in cost and required production time [11, 43–46]. However, the initial impression of the validation shows the developed system as non-applicable due to classification accuracies on the generalized task of 90–94%. For the application of inline quality monitoring, the defect detection rate is vastly higher than the classification accuracy, resulting from the repeated evaluation of data in multiple contexts due to the overlapping segments. As most misclassified segments are in defect transition zones and only individual segments of welds are incorrectly classified, the average defect detection rate of all iterations ($n=6$) exceeds 99.9% for 3 of 4 network configurations (4: 99.87%). The number of occurrences when three consecutive segments of the same weld are misclassified as false negative (defect not detected) is given in Table 9; additionally, the number of three consecutive false positives (ROI where no defect is present) is given in brackets.

When evaluating 594 weld seams in 2970 segments (6 iterations×99 weld seams×5 segments), the configuration with the lowest defect detection rate, K32, misses 4 defects. These are spread out over welds in three different alloys. The best network configuration of the training dataset, K16T32, achieves a

Table 8 Sum of misclassified segments by 6 iterations of network configuration and number and fraction of segments from the defect transition area

$N=6$	K16T32		K16T+		K32		K32T+	
	Misclass. segments	Transition area	Misclass. segments	Transition area	Misclass. segments	Transition area	Misclass. segments	Transition area
AA2017	81	50 61.73%	69	50 72.46%	90	49 54.44%	58	43 74.14%
AA7075	82	74 90.24%	61	55 90.16%	97	86 88.66%	82	72 87.80%
AA5083	5	1 20%	2	1 50%	13	9 69.23%	7	4 57.14%
AA5754	94	83 88.30%	72	56 77.78%	83	69 83.13%	76	70 92.11%
Total no./Avg. Trans	262	208 79.39%	204	162 79.41%	283	213 75.27%	223	189 84.75%

Table 9 Total number of three consecutive false negative (false positive) classified segments for $n=6$ iterations

$N=6$	K16T32	K16T+	K32	K32T+
No. segments, AA2017 1 mm	0 (3*)	2* (0)	2 (2)	0 (1)
No. segments, AA7075 1.5 mm	3* (0)	0 (0)	1 (3*)	1 (1)
No. segments, AA5083 2 mm	0 (0)	0 (0)	1 (0)	0 (0)
No. segments, AA5754 1.5 mm	0 (3*)	0 (3**)	0 (3*)	0 (2)
Average Detection Accuracy (%)	99.93	99.96	99.87	99.96

*2 × same 3 segments

**All 3 × same 3 segments

detection rate of 99.93% missing 2 defects, with one of them being missed by two different iterations and one of those iterations also accounting for the third miss, resulting in 4 of 6 iterations detecting all defects. K16T+and K32T+, the wider configurations of the respective network growth rate, both achieve defect detection accuracies of 99.96%, with the former missing the same defect twice. Thereby, 5 of 6 iterations of the widest network configuration, K32T+, accurately detect all defects and do not pass any defective welds through their quality inspection.

At the shown defect detection rates, the developed system again surpasses all current industrial quality control for inner volumetric defects in FSW in terms of detectable defect size and detection accuracy. Furthermore, the system can be implemented on edge devices at the machine level and used as inline quality monitoring with ROI determination for further local, intensified quality investigation or defect localization. The recorded process data, amended with meta-data relating to the weld, its workpieces and parameters, and localized quality information can fulfill the customer requirements of production traceability. In a next step, the system can be updated in the context of a fully connected production workshop to not only deliver localized quality data to downstream processes, but also utilize data from upstream processes to adjust the utilized meta-data to local workpiece properties and up-to-date tool and machine quality information.

5 Conclusions

Within this work, convolutional neural networks based on DenseNet201 were modified and amended with weld and workpiece meta-data. The networks were trained to classify FSW weld data in order to detect and localize inner volumetric defects (tunnel defects, voids). For a training dataset of 1405 weld seam segments from 3 different alloys, 2 sheet thicknesses, and 13 weld parameter combinations, 4 different network configurations and hyperparameter combinations were found that resulted in test set (846 segments) classification accuracies > 99%. The classification accuracies of the trained networks, as well as their ability to generalize over new parameter combinations within the training data and extrapolate beyond the training data, were validated. A validation dataset of 495 weld segments, executed on a different welding machine, including training set and new weld parameter combinations and alloys inside and outside the training parameter range, was used. An average validation accuracy of ~ 92% was achieved, with wider network configurations that spread the weights over a larger number of detectable features, outperforming narrower networks that emphasize a lower number of detection features, which provided better accuracy in classifying the training data. The major factor for misclassification was identified as the defect detection threshold, which is generalized for all welding conditions, resulting in defect initiation locations and defect end locations being misclassified at rates exceeding the average. Misclassifications of the defect initiation or defect end location account for 79.71% of the misclassifications. Furthermore, the machine and weld tool (geometry) influence, that are not yet part of the network meta-data enhancement, showed significant influence on the weld process dynamics and thereby the resulting weld data classification.

Despite the resulting classification accuracy below industrial quality monitoring application levels, the developed system already vastly outperforms the state of the art, due to the overlapping segmentation of the recorded data and the significantly reduced defect detection size of 0.063 mm for 1-mm and 1.5-mm thickness plates and 0.08-mm for 3-mm plates. The multi-contextual data evaluation of the overlapping segments results in a defect detection accuracy of 100% for all network configurations when classifying the reserved test sets of the training dataset and defect detection rates of 99.87–99.96% for the classification of the validation set data. For the best network configuration, 5 of 6 iterations detect 100% of the defects in 495 weld segments of the validation dataset, while the last iteration misses to flag one defect.

The developed system can be deployed on edge devices as an inline quality monitoring tool to determine critical ROI for further inspection, as inferencing is about 2/100 of the duration between new segments. It can also provide a subsequent evaluation step for quality determination and process traceability by recording process dynamics and amending joining task and workpiece meta-data.

Funding Open Access funding enabled and organized by Projekt DEAL. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2023 Internet of Production – 390621612.

Declarations

Competing interests The authors declare no competing interests.

References

1. Brecher C, Schuh G, van der Aalst W, Jarke M, Piller FT, Padberg M. Internet of Production. Cham: Springer International Publishing; 2024. ISBN: 978-3-031-44496-8 . https://doi.org/10.1007/978-3-031-44497-5
2. Rabe P, Reisgen U, Schiebahn A (2023) Non-destructive evaluation of the friction stir welding process, generalizing a deep neural defect detection network to identify internal weld defects across different aluminum alloys. Weld World 67(3):549–560. https://doi.org/10.1007/s40194-022-01441-y
3. Rabe P, Schiebahn A, Reisgen U (2021) Force feedback-based quality monitoring of the friction stir welding process utilizing an analytic algorithm. Weld World 65(5):845–854. https://doi.org/10.1007/s40194-020-01044-5
4. Reisgen U, Schiebahn A, Sharma R, Maslennikov A, Rabe P, Erofeev V (2020) A method for evaluating dynamic viscosity of alloys during friction stir welding. J Adv Join Process 1:100002. https://doi.org/10.1016/j.jajp.2019.100002
5. Rabe P, Schiebahn A, Reisgen U (2022) Deep learning approaches for force feedback based void defect detection in friction stir welding. J Adv Join Process 5:100087. https://doi.org/10.1016/j.jajp.2021.100087
6. Rabe P, Schiebahn A, Reisgen U. Volumetric defect detection in friction stir welding through convolutional neural networks generalized across multiple aluminum-alloys and sheet thicknesses. In: Proceedings in Engineering Mechanics. p. 43–61 https://doi.org/10.1007/978-3-031-54732-4_4
7. Thomas WM Improvements relating to friction welding. European Patent Specifications 0615 48 B1
8. Lohwasser D, Chen Z, editors. Friction stir welding: from basics to applications. Bocan Raton, Fla., Oxford: CRC Press; WP Woodhead Publ; 2010. (Woodhead Publishing in materials). ISBN: 9781845694500
9. Richter B. Robot-based friction stir welding for E-mobility and general applications. eBIS 2017; 2017(5):103–10. https://doi.org/10.17729/ebis.2017.5/11
10. Sharma N, Khan ZA, Siddiquee AN (2017) Friction stir welding of aluminum to copper—an overview. Transact Nonferrous Metals Soc Chin 27:2113–2136. https://doi.org/10.1016/S1003-6326(17)60238-3
11. Taheri H, Kilpatrick M, Norvalls M, Harper WJ, Koester LW, Bigelow T et al (2019) Investigation of nondestructive testing methods for friction stir welding. Metals 9(6):624. https://doi.org/10.3390/met9060624
12. Luhn T. Prozessdiagnose und Prozessüberwachung beim Rührreibschweißen [Zugl.: Ilmenau, Techn. Univ., Diss., 2012]. 1. Aufl. Berlin: Pro Business; 2013. ISBN: 9783863869632
13. Boldsaikhan E, Logar AM, Corwin EM. Real-time quality monitoring in friction stir welding: the use of feedback forces for nondestructive evaluation of friction stir welding. Saarbrücken: Lambert Academic Publishing; 2010. ISBN: 9783838352985
14. Mishra D, Roy RB, Dutta S, Pal SK, Chakravarty D. A review on sensor based monitoring and control of friction stir welding process and a roadmap to Industry 4.0. Journal of Manufacturing Processes 2018; 36:373–97. https://doi.org/10.1016/j.jmapro.2018.10.016
15. Das B, Pal S, Bag S (2016) A combined wavelet packet and Hilbert-Huang transform for defect detection and modelling of weld strength in friction stir welding process. J Manuf Process 22:260–268. https://doi.org/10.1016/j.jmapro.2016.04.002
16. Huang G, Liu Z, van der Maaten L, Weinberger KQ. Densely connected convolutional networks. 1063-6919 2017:2261–9. https://doi.org/10.1109/CVPR.2017.243
17. Arbegast WJ. Application of friction stir welding and related technologies: Kapitel 13. In: Mahoney, M. W., Mishra, R.S. (Eds.): Friction stir welding and processing. Materials Park, OH, USA 2007
18. Mishra RS. Friction stir welding and processing: science and engineering. 1st ed. Cham: Springer International Publishing AG; 2014. ISBN: 9783319070438
19. Mishra RS, De PS, Kumar N. Friction stir welding and processing: science and engineering. Cham, Heidelberg: Springer; 2014. ISBN: 9783319070421 . https://doi.org/10.1007/978-3-319-07043-8
20. Gebhard P. Dynamisches Verhalten von Werkzeugmaschinen bei Anwendung für das Rührreibschweißen [Zugl.: München, Techn. Univ., Diss., 2010]. München: Utz; 2011. (Forschungsberichte / IWB; vol 253). ISBN: 9783831641291
21. Hattingh DG, Blignault C, van Niekerk TI, James MN (2008) Characterization of the influences of FSW tool geometry on welding forces and weld tensile strength using an instrumented tool. J Mater Process Technol 203(1–3):46–57. https://doi.org/10.1016/j.jmatprotec.2007.10.028
22. Hasieber M, Wenz F, Grätzel M, Lenard JA, Matthes S, Bergmann JP (2023) A systematic analysis of maximum tolerable tool wear in friction stir welding. Weld World 67(2):325–339. https://doi.org/10.1007/s40194-022-01407-0
23. Cole EG, Fehrenbacher A, Shultz EF, Smith CB, Ferrier NJ, Zinn MR et al (2012) Stability of the friction stir welding process in presence of workpiece mating variations. Int J Adv Manuf Technol 63(5–8):583–593. https://doi.org/10.1007/s00170-012-3946-1
24. Więckowski W, Burek R, Lacki P, Łogin W. Analysis of wear of tools made of 1.2344 steel and MP159 alloy in the process of friction stir welding (FSW) of 7075 T6 aluminium alloy sheet metal. EiN 2018; 21(1):54–9. https://doi.org/10.17531/ein.2019.1.7
25. Muhayat N, Zubaydi A, Sulistijono, Yuliadi MZ. Effect of tool tilt angle and tool plunge depth on mechanical properties of friction stir welded AA 5083 joints. Advances in Applied Mechanics and Materials 2014; 493:709–14. https://doi.org/10.4028/www.scientific.net/AMM.493.709
26. Zettler R, Lomolino S, dos Santos JF, Donath T, Beckmann F, Lippman T et al (2005) Effect of tool geometry and process parameters on material flow in FSW of an AA 2024–T351 alloy. Weld World 49(3–4):41–46. https://doi.org/10.1007/BF03266474
27. Fehrenbacher A, Duffie NA, Ferrier NJ, Pfefferkorn FE, Zinn MR (2014) Effects of tool–workpiece interface temperature on weld quality and quality improvements through temperature control in friction stir welding. Int J Adv Manuf Technol 71(1–4):165–179. https://doi.org/10.1007/s00170-013-5364-4

28. Franke D, Rudraraju S, Zinn M, Pfefferkorn FE (2020) Understanding process force transients with application towards defect detection during friction stir welding of aluminum alloys. J Manuf Process 54:251–261. https://doi.org/10.1016/j.jmapro.2020.03.003
29. Jene T. Entwicklung eines Verfahrens zur prozessintegrierten Prüfung von Rührreibschweißverbindungen des Leichtbaus sowie Charakterisierung des Ermüdungsverhaltens der Fügungen [Zugl.: Kaiserslautern, Techn. Univ., Diss., 2008]. Als Ms. gedr. Kaiserslautern: Techn. Univ. Lehrstuhl für Werkstoffkunde; 2008. (Werkstoffkundliche Berichte; vol 21). ISBN: 3932066200
30. Roberts J. Weld quality classification from sensory signatures in friction-stir-welding (FSW) using discrete wavelet transform and advanced metaheuristic techniques. LSU Master's Theses 2016
31. Hattingh DG, van Niekerk TI, Blignault C, Kruger G, James MN (2004) Analysis of the FSW force footprint and its relationship with process parameters to optimise weld performance and tool design. Weld World 48(1–2):50–58. https://doi.org/10.1007/BF03266414
32. Enkhsaikhan Boldsaikhan, Edward M. Corwin, Antonette Logar, William J. Arbegast. Neural network evaluation of weld quality using FSW feedback data
33. Boldsaikhan E, Corwin EM, Logar AM, Arbegast WJ (2011) The use of neural network and discrete Fourier transform for real-time evaluation of friction stir welding. Appl Soft Comput 11(8):4839–4846. https://doi.org/10.1016/j.asoc.2011.06.017
34. Hartl R, Bachmann A, Habedank JB, Semm T, Zaeh MF (2021) Process monitoring in friction stir welding using convolutional neural networks. Metals 11(4):535. https://doi.org/10.3390/met11040535
35. Wei J. AlexNet: the architecture that challenged CNNs. Towards Data Science 2019
36. Rabe P, Motschke T, Schiebahn A, Reisgen U. Methode zur Umsetzung von Rührreibschweißprozessen auf konventionellen Fräsmaschinen mittels eines empirischen Ansatzes. Schweissen und Schneiden 2020; 72(1–2)
37. Ambrosio D, Wagner V, Dessein G, Paris J-Y, Jlaiel K, Cahuc O (2021) Plastic behavior-dependent weldability of heat-treatable aluminum alloys in friction stir welding. Int J Adv Manuf Technol 117(1–2):635–652. https://doi.org/10.1007/s00170-021-07754-4
38. Kerckhofs G, Schrooten J, van Cleynenbreugel T, Lomov SV, Wevers M (2008) Validation of X-ray microfocus computed tomography as an imaging tool for porous structures. Rev Sci Instrum 79(1):13711. https://doi.org/10.1063/1.2838584
39. Viscom. X-ray tubes; 2021 [cited 2022 Nov 16]
40. International Organization for Standardization. Non-destructive testing — image quality of radiographs: determination of the image unsharpness and basic spatial resolution value using duplex wire-type image quality indicators. 2018th ed.
41. Wuest T, Weimer D, Irgens C, Thoben K-D (2016) Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research 4(1):23–45. https://doi.org/10.1080/21693277.2016.1192517
42. Alexander Amini. Introduction to deep learning: MIT Course 2021 [cited 2021 Dec 13]
43. See JE. Visual Inspection: A Review of the Literature. Sandia National Laboratories 2012
44. Spencer F, United States. Dept. of Transportation, Sandia National Laboratories. Visual inspection research project report on benchmark inspections; 1996 DOT/FAA/AR-96/65. https://doi.org/10.21949/1403546
45. Burford DA, Gimenez Britos P., Boldsaikhan E, Brown J. Evaluation of friction stir weld process and properties for aerospace application: e-NDE for friction stir processes. FAA Joint Advanced Materials & Structures (JAMS) Center of Excellence 2010
46. Huggett D. Friction stir welding manufacturing advancement by on-line high temperature phased array ultrasonic testing and correlation of process parameters to joint quality [Luisiana State University, Doctoral Dissertation]; 2017. https://doi.org/10.31390/gradschool_dissertations.4139