





Resolution enhancement of scanning electron micrographs using artificial intelligence

T. Reclik^{a,*,}, S. Medghalchi^a, P. Schumacher^a, M.A. Wollenweber^{a, }, T. Al-Samman^{a, },
S. Korte-Kerzel^{a, }, U. Kerzel^{b, }

^a Institute for Physical Metallurgy and Materials Physics, RWTH Aachen University, Aachen, Germany

^b Data Science and Artificial Intelligence in Materials and Geoscience, Fakultät für Georessourcen und Materialtechnik, RWTH Aachen University, Aachen, Germany

ARTICLE INFO

Keywords:

Scanning electron microscopy
Artificial intelligence
Computer vision
Deep learning
Resolution enhancement
Texture transformer super-resolution network
Reference-based super-resolution
Dual-phase steel
Case-hardening steel
16MnCrS5

ABSTRACT

Microscopic imaging of materials often requires the examination of large sample areas at high magnification to identify and analyse rare structural features. High-resolution imaging in scanning electron microscopy is particularly time-intensive, as images are acquired through sequential scanning of an electron beam across the sample surface. This presents a critical challenge, as researchers must balance imaging speed and resolution while ensuring statistically meaningful observations of sparsely distributed features. To address these challenges, we present a novel resolution enhancement method for electron microscopy based on artificial intelligence. Suitable reference images are selected using vector embeddings and processed by a texture-transformer network. Using a tailored dataset of dual-phase steel micrographs, we demonstrate that our trained network outperforms traditional interpolation methods in both quantitative similarity metrics and crucial material-specific features, such as phase boundaries and microstructural voids. The method's transferability is validated using micrographs from a 16MnCrS5 case-hardening steel sample. We achieve a 16-fold acceleration through resolution trade-offs between target and recording and additionally propose a scan-enhance-rescan workflow where resolution-enhanced micrographs guide the identification of regions of interest for targeted high-resolution rescanning. We provide quantitative estimates of expected time savings, offering a practical framework for efficient high-resolution microscopy across large areas.

1. Introduction

Scanning electron microscopy (SEM) is essential for understanding and optimizing metallic materials, yet the examination of large sample areas at high magnification presents a significant challenge. This is particularly important where rare microstructural features at the sub-micron level govern the material's properties at the macroscopic scale. This is the case, for example, in many steels, where manufacturing processes induce damage from deformation steps. Understanding this deformation induced damage as rare events analysed by large area but high resolution microscopy as a basis to design and control better processing strategies of semi-finished products, such as sheet metal, may have a significant effect on the amount of material that has to be used to guarantee safety in any final product. Using less material is a common strategy to mitigate environmental impact. Steel production, being a major contributor to anthropogenic greenhouse gas emissions [1,2], accounts for approximately a quarter of global industrial CO₂ emissions in

2019 [3]. Therefore, efficient microscopic characterisation methods are crucial for developing optimized microstructures that can improve material properties while reducing energy consumption during production.

To achieve these optimisations, researchers must address several key challenges in understanding and controlling material behaviour during manufacturing and service and the methodological requirements for generating this understanding from experimental data. One related challenge is the tendency of multi-phase alloys to form damage sites in the shape of voids during forming operations at the interfaces of phases or through the fracture of one phase [4]. By understanding damage nucleation and evolution, the active formation mechanisms and how they relate to the underlying multi-phase microstructures, processing routes can be adjusted to improve material performance and allow for more precise determination of component safety and lifetimes [5,6]. Scanning electron microscopy offers unique capabilities to resolve the underlying microstructure, including damage sites, at high resolution. This allows to understand the relationship between the behaviour of these intricate

* Corresponding author.

E-mail address: reclik@imm.rwth-aachen.de (T. Reclik).

<https://doi.org/10.1016/j.matdes.2025.113955>

Received 13 February 2025; Received in revised form 28 March 2025; Accepted 11 April 2025

Available online 16 April 2025

0264-1275/© 2025 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

structures and physical mechanisms. However, since damage formation and evolution are inherently stochastic processes, electron micrographs must cover large areas at high resolution to be able to draw statistically meaningful conclusions [7–10].

The analysis of the resulting increasingly large datasets through manual evaluation is labour intensive. Therefore, in previous studies, tools have been developed to automate data evaluation [7,9,11]. These tools either combine traditional computer vision techniques with deep learning methods [7,11] or employ a full deep learning based pipeline [9]. While these tools enable the automated analysis of large datasets, the acquisition of large areas still requires long imaging times, posing as a bottleneck in the study of damage. Recent developments in deep learning-based image enhancement, especially in consumer photography [12,13], medical imaging [14–16], and first applications to material science [17] suggest untapped potential for time-efficient imaging. These deep learning-based super-resolution algorithms have emerged as powerful techniques for enhancing image quality. At its core, super-resolution aims to reconstruct high-resolution images from low-resolution counterparts, effectively increasing pixel density while preserving or recovering fine details. The two main approaches are single-image super-resolution [18,19], i.e. the reconstruction of high-resolution information from its low-resolution counterpart, and reference-based super-resolution [12,13], where high-resolution reference image are additionally supplied.

Building on these advances in super-resolution techniques, we propose a novel approach to accelerate large-area imaging in scanning electron microscopy. Conventionally, large areas are imaged at resolutions required for the subsequent analysis of damage sites. We hypothesise that by using a fraction of the full resolution combined with the successive application of deep learning-based super-resolution techniques, imaging times can be significantly reduced. This approach enables or greatly expands the scope of experiments that aim to understand and control deformation-induced damage and accelerate the evaluation of a broader range of parameter combinations and processing routes, for example by large-area in-situ experiments to observe the formation and evolution of damage sites [20,9] or large-area slice-and-view experiments for comprehensive three-dimensional analysis [21]. Ultimately, the approach can also be transferred to many other investigations of features or processes in multiphase microstructures of metallic or other solid materials to enhance the efficiency of material characterisation and process optimisation.

In this work, we implement a pipeline for the automated selection of reference images, train and evaluate a texture transformer network for super-resolution (TTSR) [12] on two materials, a dual-phase steel – as a proof of concept – and a 16MnCrS5 case-hardening steel – to study the method's transferability. These materials have different microstructures and are both studied for their damage behaviour [7,11,21,9].

2. Methods

2.1. Materials

Both dual-phase and case-hardening steels are pivotal in numerous industrial applications due to their distinct properties and widespread usage in sectors such as automotive manufacturing and heavy machinery. We used a commercial DP800 dual phase steel in sheet form (ThyssenKrupp Steel Europe AG). Damage in these very different microstructures occurs in characteristic locations and by different underlying mechanisms, which we describe briefly below and refer to by the short keywords given in brackets for each case. Dual-phase steels consist of a comparatively brittle martensite phase and a ductile ferrite phase, shown in Fig. 1 (a). During forming operations damage sites nucleate in the form of microscopic voids at the interface of martensite and ferrite (interface decohesion), at the grain boundary of two ferrite grains (boundary decohesion) or through the cracking of martensite islands (martensite crack) [22]. 16MnCrS5 case-hardening steel consists

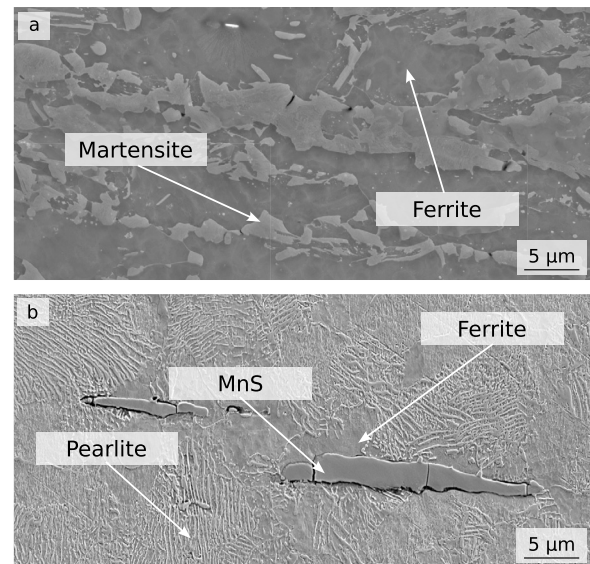


Fig. 1. Labelled electron micrographs of (a) a commercial DP800 dual-phase steel and (b) 16MnCrS5 case-hardening steel.

of pearlite and ferrite with dispersed manganese sulfide (MnS) inclusions, illustrated in Fig. 1 (b). We use here a commercial case-hardening steel sample (Georgsmarienhütte Holding GmbH, Germany). Damage in this material tends to form at the interface of the MnS inclusion to the surrounding steel matrix (decohesion) or through the cracking of MnS inclusions (crack). The microstructures of both materials after metallographic preparation are visually different when imaged in the SEM, as can be seen in Fig. 1.

2.1.1. Sample preparation

A bending sample (dual-phase steel) and a tensile test sample (16MnCrS5 case-hardening steel) were cut using an electrical discharge machine. Then, both samples underwent a similar metallographic preparation process: in a first step, the surface of both samples was ground using sandpaper with grits ranging from 800 to 4000, using water as a cooling agent. Subsequently, the sample surface was mechanically polished using 6 µm, 3 µm and 1 µm water-based diamond suspensions. In this step, a DAC cloth was used for the dual-phase steel sample, while an Alpha cloth was used for the 16MnCrS5 case-hardening steel sample. Both samples were finished by polishing with a 0.25 µm Oxide Polishing Suspension (OPS) for 1 min. In a last step, the samples were etched in Nital solution to achieve a topographical contrast, allowing to discriminate between the different phases of the samples under the scanning electron microscope. The dual-phase steel was etched in 1% Nital solution for 5 s, while the case-hardening steel was etched in 5% Nital solution for 5 s.

Both samples were deformed in the same microMECHA Proxima testing stage using the appropriate module for the respective sample geometry. The dual-phase steel bending sample was deformed using the three-point bending test module up to a plastic strain of 11% at the outer edge of the sample. This deformation step took place before metallographic preparation of the sample. In contrast, the tensile test of the case-hardening steel sample was conducted after metallographic preparation, and terminated at a plastic strain of 3%. After preparation and deformation the samples were mounted in the SEM chamber for the analysis of the prepared surface.

2.2. Image acquisition

Supervised deep learning algorithms require both target and input data to learn. For deep-learning super-resolution algorithms, these correspond to the target high-resolution images and low-resolution counter-

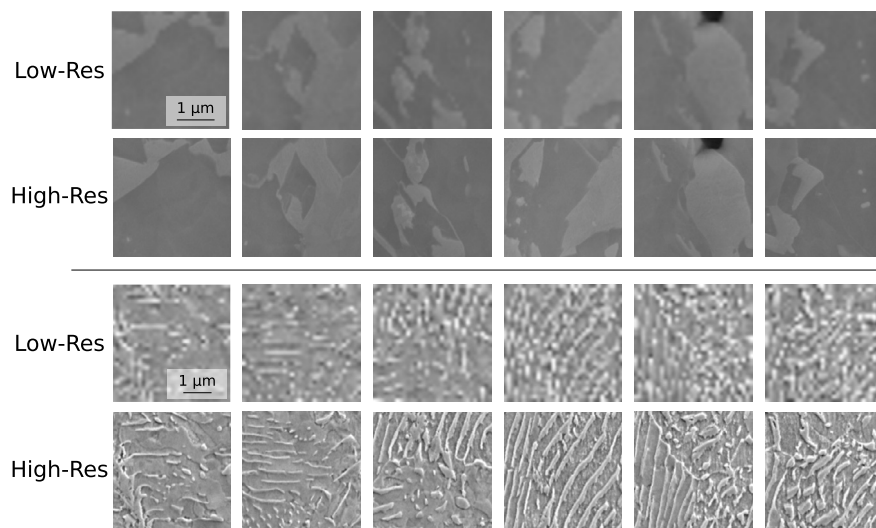


Fig. 2. Examples for pairs of high-resolution and low-resolution images used as training data. Top: dual-phase steel, bottom: 16MnCrS5 case-hardening steel.

parts. In the case of reference-based super-resolution algorithms, a stack of reference images is additionally required. Here, the target resolution was chosen to equal the resolution used in our previous studies of damage [7,9]. As the pre-trained network weights are trained to increase the resolution by a factor of four [12], the lower resolution corresponds to a quarter of the target resolution for both cases. Furthermore, the input size is constrained to 32×32 pixel, the output size to 128×128 pixel and the high-resolution reference images to 300×300 pixel.

Deep learning-based super-resolution algorithms follow two approaches to acquire corresponding low-resolution counterparts to their target images. Either low-resolution images are generated from the target images through interpolation [19,18,12] or they are both recorded using special devices [23]. In consumer photography the latter approach has additional challenges, as the resolution cannot be seen as an independent parameter. Different resolutions result from physically different cameras which differ also in field of view, curvature, and colour distributions. Therefore most often, the low-resolution images are generated from high-resolution images by image processing rather than recorded separately. In the SEM however, due to the sequential acquisition of data points, recording the same field of view in two different resolutions poses fewer challenges. Because of the ease of acquisition and to prevent possible noise transfer from high-resolution images to generated low-resolution images, we recorded both high and low-resolution images. These images then need to be aligned afterwards, as effects such as drift can influence the position of features. These effects accumulate with recording time, therefore the recording window size has a strong influence on the data quality. In this work we recorded three different datasets, the first consists of 25,000 image pairs on the dual-phase steel sample, recorded at the lower bound of possible window sizes, given by the target size of the reference-based super-resolution algorithm, i.e. 128×128 for the high-resolution images and 32×32 for the low-resolution images, representing a field of view of $3.125 \mu\text{m}$. Then additionally, one larger pair of micrographs at the other extreme, determined by the microscope interface, here 4096×4096 for the high-resolution images and 1024×1024 for the low-resolution images, both corresponding to a field of view of $100 \mu\text{m}$, was recorded on the same sample, to investigate its performance when generating larger images and give researchers more contextual information. Lastly, to investigate the transferability of this approach, 160 image pairs with window sizes 3072×3072 and 976×976 , again representing a field of view of $100 \mu\text{m}$ were recorded on the 16MnCrS5 sample, of which a fraction was used here. A summary of datasets can be found in Table 1 and a few examples of subdivided and aligned images can be seen in Fig. 2.

The electron micrographs were acquired using a TESCAN CLARA (Tescan Group, Czech Republic) scanning electron microscope using secondary electrons for detection. For the dual-phase steel sample, an accelerating voltage of 20 kV and a beam current of 3 nA was used. For the 16MnCrS5 case-hardening steel sample, an accelerating voltage of 10 kV and a beam current of 1 nA was used.

2.3. Reference-based super-resolution

Reference based super-resolution aims to reconstruct high-resolution information given a low-resolution input and a suitable additional high-resolution reference image. Since super-resolution is mathematically ill-posed, with multiple possible high-resolution outputs mapping to the same low-resolution input, the additional high-resolution reference image can provide constraints for the reconstruction process. To implement this approach, we adopted the texture transformer architecture from [12], leveraging their pre-trained network weights through transfer learning. Our implementation follows a two-step process: first, a reference selector assigns appropriate reference images to each input image. These paired low-resolution and reference image batches are then processed by the texture transformer network to produce the resolution-enhanced output. The complete workflow is illustrated in Fig. 3. Since the network was originally trained on colour images, we transform our grey scale SEM micrographs to the expected colour format and normalize these channels to fall within the interval $[-1, 1]$.

2.3.1. Reference selector

The reference selector identifies reference images with the highest similarity to the low-resolution input. Central to this process is the calculation of embedding vectors using a truncated convolutional neural network, specifically the VGG19 network [24]. We truncated the VGG19 layers in a manner consistent with the subsequent texture transformer network after the 13th convolutional layer and initialized it with weights pre-trained on the ImageNet dataset for image classification [25]. The selection process can be formalised as follows:

1. Transform the low-resolution input images and the high-resolution reference images to the same image domain: using bi-cubic interpolation, up-sample the low-resolution image to high-resolution. To be able to compare the similarity between the high-resolution reference images and the low-resolution input, remove high-resolution information from the reference images through subsequent down- and up-sampling.

Table 1

Overview of recorded datasets. The data in the “DP-small” data for the dual-phase steel were used for the training of the neural networks, whereas the single image in the “DP-large” dataset was only recorded for inference and has therefore no training and validation data.

Name	DP-small	DP-large	16MnCrS5
Recorded image pairs	25 000	1	160
Low-resolution window size	32×32	1024×1024	768×768
High-resolution window size	128×128	4096×4096	3072×3072
Low-resolution	98 nm/px	98 nm/px	102 nm/px
High-resolution	24 nm/px	24 nm/px	33 nm/px
Dwell time	100 μ s	100 μ s	32 μ s
Recorded reference images	256	-	-
Window size	300×300	-	-
Reference resolution	24 nm/px	-	-
Dwell time	100 μ s	-	-
Number of training pairs	20 000	-	1600
Number of validation pairs	2500	-	200
Number of test pairs	2500	1700	200
Number of reference images	256	256 (Same as DP-small)	256

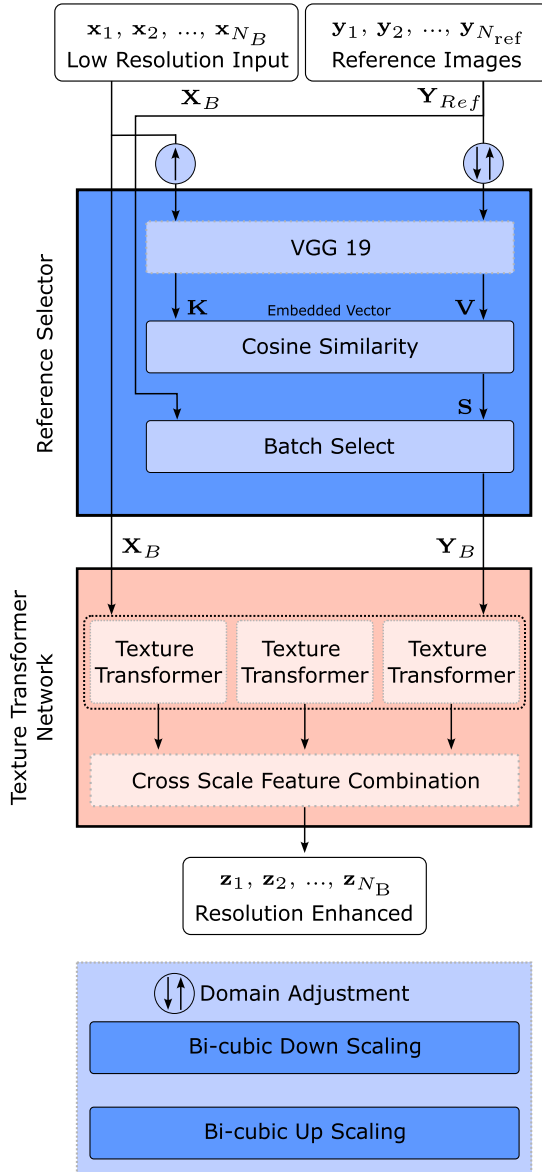


Fig. 3. Two step workflow consisting of a reference selector and a texture transformer network.

2. Calculate embedding vectors \mathbf{K} and \mathbf{V} for the low-resolution and reference images using the truncated VGG19 network

$$\mathbf{K} = \text{VGG19}(\mathbf{X}_B), \quad (1)$$

$$\mathbf{V} = \text{VGG19}(\mathbf{Y}_{\text{Ref}}) \quad (2)$$

3. Using the resulting embedding vectors calculate the cosine similarity between reference and low-resolution images:

$$S_{ij} = \frac{(\mathbf{K} \cdot \mathbf{V}^T)_{ij}}{\sqrt{\sum_k K_{ik}^2} \sqrt{\sum_l V_{lj}^2}} \quad (3)$$

4. Select the indices of the most similar reference images:

$$B_j = \arg \max_i (S_{ij}) \quad (4)$$

5. Construct the batch of selected reference images:

$$\mathbf{Y}_{B,i} = \mathbf{Y}_{\text{Ref}, B_i} \quad (5)$$

To optimize efficiency during training, we pre-determined the highest similarity reference image for each low-resolution image. Post-training, the weights of the trained texture extractor from the texture transformer network [12] can be used to initialize the VGG19 network in the reference selector, potentially enhancing its selection capabilities.

2.3.2. Resolution enhancement using texture transformers

The texture transformer network for super-resolution combines three major concepts. First the extraction of texture using a truncated VGG19 network [24]. Texture here refers to the intermediate representation of image information in convolutional neural networks. Using this texture information, an attention mechanism [26] aligns features from the reference image with the input and then reconstructs the super-resolved output. These so-called texture transformers aim to improve the resolution at different scales. To achieve a four-fold resolution increase, intermediate outputs at single-fold, two-fold, and four-fold are combined in a network [12]. During training this output is compared to the high-resolution ground truth and weights are adjusted according to the loss function. Here we followed the same approach as in [12] and used a combination of the following loss functions.

- **Reconstruction Loss:** The reconstruction loss calculates the pixel-wise difference between output and ground truth and is defined as

$$\mathcal{L}_{\text{rec}} = \frac{1}{CHW} \|I^{HR} - I^{SR}\|_1 \quad (6)$$

Here, C , H , and W are the number of colour channels, the height, and the width, I represents the image, and $\|\dots\|_1$ is the L_1 norm of the difference. Only using this loss function generally results in overly smooth predictions, as a conservative prediction will on average result in a lower loss value.

- **Perceptual Loss:** The perceptual loss calculates the difference in an intermediate layer of a trained neural network between the generated enhanced image and the ground truth [27] and has been specifically developed for the application of neural networks in resolution-enhancement algorithms.

$$\mathcal{L}_{per} = \frac{1}{C_i H_i W_i} \|\phi_i^{vgg}(I^{SR}) - \phi_i^{vgg}(I^{HR})\|_2^2 + \sum_j \frac{1}{C_j H_j W_j} \|\phi_j^{lte}(I^{SR}) - T\|_2^2 \quad (7)$$

ϕ_i^{vgg} is the VGG19 network truncated after the i th layer, C_i , H_i , and W_i are the number of channels, the height and width of that layer. Similarly ϕ_j^{lte} is the learnable texture extractor truncated after the j th layer, C_j , H_j , and W_j are the number of channels, the height and width of the j th layer. Here we chose i to be the layer after the 13th convolutional layer and sum over j to be chosen after the first convolutional layer, the third convolutional layer, and after the fifth convolutional layer, following again the approach in [12].

- **Adversarial Loss:** Finally, the adversarial loss is based on the idea behind generative adversarial networks (GANs) [28]. Adversarial networks typically consist of two competing networks, where the generator network creates new images and the discriminator network tries to tell the output from the generator apart from authentic images. As done in [12], we chose the Wasserstein loss [29] given by:

$$\mathcal{L}_{adv} = \mathbb{E}_{\tilde{x} \sim \mathbb{P}_g} [D(\tilde{x})] - \mathbb{E}_{x \sim \mathbb{P}_d} [D(x)] + \lambda \mathbb{E}_{\tilde{x} \sim \mathbb{P}_{\tilde{x}}} \left[\left(\|\nabla_{\tilde{x}} D(\tilde{x})\|_2 - 1 \right)^2 \right] \quad (8)$$

where D is the discriminator, \mathbb{E} is the expectation value over \mathbb{P}_g , the distribution over generated images after the resolution-enhancement process, \mathbb{P}_d the distribution of ground truth images, $\mathbb{P}_{\tilde{x}}$, the distribution of all possible images, λ is a regularisation parameter that controls the degree of gradient penalisation. Following [12] we used $\lambda = 10$.

Combining all parts, the final loss function is a weighted combination of the three contributions:

$$\mathcal{L}_{tot} = \lambda_{rec} \mathcal{L}_{rec} + \lambda_{per} \mathcal{L}_{per} + \lambda_{adv} \mathcal{L}_{adv} \quad (9)$$

Training with the full loss function from the beginning of the training can lead to unstable behaviour, where the competing contributions can produce undesired outputs. This is well known for the case of generative adversarial networks [30]. To mitigate this behaviour, we follow the approach by Yang et al. [12] and train the network for 80 epochs, the first 40 epochs only with the reconstruction loss and then add the other contributions later for the last 40 epochs.

2.3.3. Model training

We divided each dataset into 80% training data, 10% validation data, and 10% test data. This corresponds to 20000, 2500, and 2500 high-resolution, low-resolution, reference image triplets, respectively for the first dual-phase steel dataset. As we recorded images in a larger window size than the network can process at once for both the dual-phase steel and the 16MnCrS5 sample, we first needed to divide the larger images into patches. Effects such as drift play an important role for these larger image sizes, therefore the low- and high-resolution image pairs needed to be realigned. We performed this alignment by defining a search window in the high-resolution image, then down-sample the high-resolution image using nearest-neighbour sampling, and calculated the position in the high-resolution image corresponding to the highest structural simi-

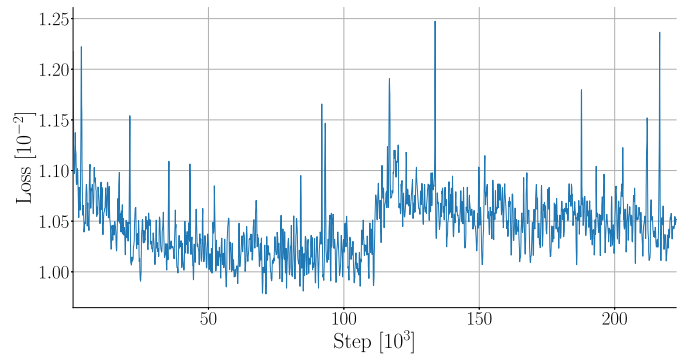


Fig. 4. Loss function during training progression. It quickly converges to a value at around 0.0112. After step 100,000 the remaining loss functions are added to the reconstruction loss, resulting in a slight increase.

ilarity score. We generated 2000 training image pairs in this manner. We also selected two images and subdivided them into 256 reference images of size 300×300 pixel. The further technical details of the training are as follows: we used a cyclic learning rate scheduler [31] together with the Adam optimiser [32]. The neural network was trained using an NVIDIA Tesla V100 SXM2 16 GB at the RWTH ITC high-performance cluster. The training took 20 h for the dual-phase steel dataset and 3 h for the 16MnCrS5 case-hardening steel dataset. As an example the loss function during training on the dual-phase steel dataset is presented in Fig. 4. In the first phase only the reconstruction loss is optimised until it starts to converge at a value around 0.0101. Then the remaining loss components were added to the total loss function and its value spiked before converging around a value of 0.0105.

2.4. Evaluation criteria

In a first step, we established a baseline to evaluate the performance of our resolution enhancement with the texture transformer network against by using conventional interpolation methods to enhance the resolution of the low-resolution electron micrographs. These interpolation methods use different families of functions, determine their parameters at positions with known values, and then evaluate them at positions of unknown values.

- **Nearest-Neighbour Interpolation:** Nearest-neighbour interpolation uses a step function, missing pixels are given the pixel value of the nearest known pixel value.
- **Bi-Cubic Interpolation:** Bi-cubic interpolation employs cubic functions in both dimensions to interpolate pixel values.
- **Lanczos Interpolation:** Lanczos interpolation [33] uses $\text{sinc}(x) = \frac{\sin(x)}{x}$ functions to calculate pixel values for interpolated pixels.

In the next step, we evaluated the quality of the images after resolution enhancement quantitatively for both the three baseline methods and the output of the texture transformer neural network using three metrics:

- **Structural Similarity Index Measure (SSIM):** The SSIM quantifies the structural integrity preserved in the images resulting from the machine-learning algorithm compared to the (high-resolution) ground-truth data. This metric compares the luminance, contrast, and structure between the reconstructed and high-resolution images. It takes values between -1 and 1 , where 1 indicates a perfect match and -1 a complete mismatch [34].
- **Peak Signal-to-Noise Ratio (PSNR):** The PSNR is, essentially, an extension of the mean squared error to images and measures the ratio between the maximum possible power of the signal, here the maximum pixel value, and the corrupting noise, the squared difference between the pixel values of the predicted image and the

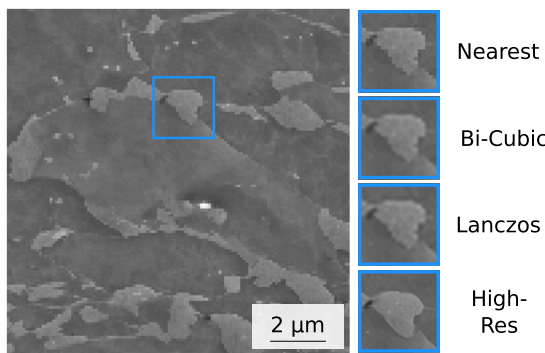


Fig. 5. Low resolution input and comparison of nearest-neighbour, bi-cubic, and Lanczos interpolation with the high-resolution ground truth of the area highlighted in blue on a dual-phase steel scanning electron micrograph.

high-resolution ground truth. The value of PSNR is expressed in dB, the higher the value the higher the similarity between the images.

- **Edge-Based Structural Similarity Index Measure (ESSIM):** While SSIM provides valuable information about overall structural similarity, it may not fully capture the quality of reconstructed fine structural details crucial in SEM imaging. ESSIM extends the standard SSIM by giving higher weight to edge regions during the similarity assessment [35], making it more sensitive to blur and edge degradation. This approach is especially relevant for evaluating super-resolution methods applied to SEM images, where edge preservation and enhancement of fine structural details are critical. Like SSIM, ESSIM produces values between -1 and 1 , with 1 indicating perfect similarity.

3. Results

3.1. Evaluation of network performance on dual-phase steel

To establish a benchmark, we first evaluated the performance of conventional interpolation methods. These results served as a baseline against which the results obtained from the texture transformer network were compared. A visual representation of these methods is provided in Fig. 5. Nearest-neighbour interpolation was observed to produce block-like artefacts, while both bi-cubic and Lanczos interpolation resulted in noticeably blurred images. While both Lanczos interpolation and bi-cubic interpolation produced visually similar results, bi-cubic interpolation outperformed Lanczos interpolation in terms of SSIM, ESSIM, and PSNR. We therefore chose bi-cubic interpolation as a visual benchmark in the following analysis.

Subsequently, we assessed the performance of the texture transformer neural network for resolution-enhancement. This evaluation was conducted using a dataset of images recorded at the network's native input resolution, as outlined in subsection 2.2. Fig. 6 shows visual examples of the low-resolution input, the expected high-resolution output that should be recovered, the result from the texture transformer network (TTSR), as well as the result from bi-cubic interpolation. The quantitative metrics are summarised in Table 2. Our analysis revealed that while conventional similarity metrics (SSIM, PSNR) showed only modest improvements, ESSIM, which specifically accounts for edge preservation, demonstrated more substantial enhancement with the TTSR approach (0.400 ± 0.006) compared to bi-cubic interpolation (0.354 ± 0.005). This quantitative edge-preservation advantage aligns with detailed visual inspection which highlighted significant differences. The neural network-based method consistently produced results closer to the expected high-resolution version, particularly in preserving fine structural details. In contrast, bi-cubic interpolation results exhibited consistent limitations, producing blurred images with diffused interfaces and softened edges. The superiority of the texture transformer network was particularly evident in reconstructing material-specific properties,

Table 2

Similarity metrics for bi-cubic and Lanczos interpolation, as well as the texture transformer neural network applied to the dual-phase steel dataset with an image size of 128×128 pixels.

Metric	Bi-Cubic	Lanczos	TTSR
SSIM	0.622 ± 0.002	0.574 ± 0.002	0.626 ± 0.002
PSNR	$25.57 \text{ dB} \pm 0.05 \text{ dB}$	$24.38 \text{ dB} \pm 0.04 \text{ dB}$	$25.96 \text{ dB} \pm 0.04 \text{ dB}$
ESSIM	0.354 ± 0.005	0.340 ± 0.005	0.400 ± 0.006

as demonstrated in Fig. 6 (a). In this example, where a brittle martensite island had cracked into two pieces, the network accurately reconstructed the separation visible in the high-resolution ground truth. The bi-cubic interpolation, however, introduced unphysical bridges between the separated pieces - an artifact that could lead to misinterpretation of the material's structural properties. Notably, the network demonstrated appropriate restraint in cases where fine structural details were ambiguous. For instance, with very fine granular martensite structures, as shown in Fig. 6 (d), the network predicted smooth structures rather than introducing potentially incorrect details. Further examples of the network's predictions compared to the high-resolution ground truth and interpolation can be found in the supplementary material section B.

In terms of similarity metrics, the texture transformer network improves over bi-cubic and Lanczos interpolation, outlined in Table 2. A direct comparison of the mean values reveals a substantial improvement between the resolution-enhanced output and the interpolated results. While the network in terms of structural similarity index only improved from 0.622 ± 0.002 to 0.626 ± 0.002 , the network improved in terms of peak-signal-to-noise ratio to a higher degree from $25.57 \text{ dB} \pm 0.05 \text{ dB}$ to $25.96 \text{ dB} \pm 0.04 \text{ dB}$. Most notably, the edge-based structural similarity index showed the most significant improvement from 0.354 ± 0.005 with bi-cubic interpolation to 0.400 ± 0.006 with the TTSR approach. This substantial 13% improvement in ESSIM underscores the network's superior ability to preserve critical edge information, which is particularly relevant for microstructural analysis.

3.1.1. Prediction of larger micrographs

To evaluate the network's capability to enhance larger micrographs, we tested it on images of 4096×4096 pixels (DP-large in Table 1). This larger field of view is crucial as microstructural features are influenced not only by their immediate surroundings but also by longer-range structural correlations.

Before these larger images could be processed by the neural network we needed to split them into smaller patches corresponding to the input size of the neural network. Following resolution-enhancement, the images were stitched to the size of the original micrograph. As artefacts at the borders of the prediction can occur, the smaller patches were generated with overlap in such a way that only the central 62.5% of the prediction were used in the stitching process without loss of information. A small section of the super-resolution micrograph after the stitching process is shown in Fig. 7. This section is 1024×512 pixels large which equals 8×4 images of the output size of the network. No stitching artefacts can be seen.

In order to compensate for accumulated drift of the electron beam during the imaging process, the low-resolution patches needed to be realigned with the high-resolution patches before calculating the similarity metrics and performing any direct visual comparison. In the evaluation of the metrics, bi-cubic interpolation was marginally outperformed by the TTSR method, which achieved the highest SSIM and PSNR values. While Lanczos interpolation delivered comparable results, it fell slightly below both bi-cubic and TTSR, further emphasising the TTSR method's effectiveness. Additionally, ESSIM showed more substantial improvement with TTSR (0.574 ± 0.005) compared to bi-cubic interpolation (0.559 ± 0.005) and Lanczos interpolation (0.546 ± 0.005), highlighting the network's ability to maintain critical edge details. The results for the similarity metrics are summarised in Table 3.

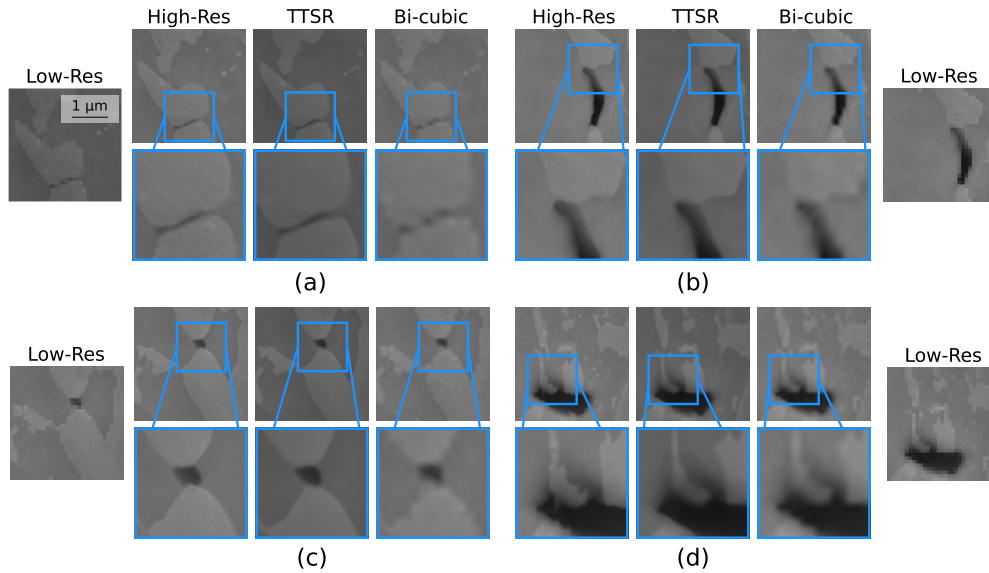


Fig. 6. Scanning electron micrographs along with the output of the texture transformer network. Each group shows the low-resolution input to the network, the high-resolution ground truth, the networks output, and for comparison bi-cubic interpolation both for the size of the entire patch as well as a zoom around each damage site.

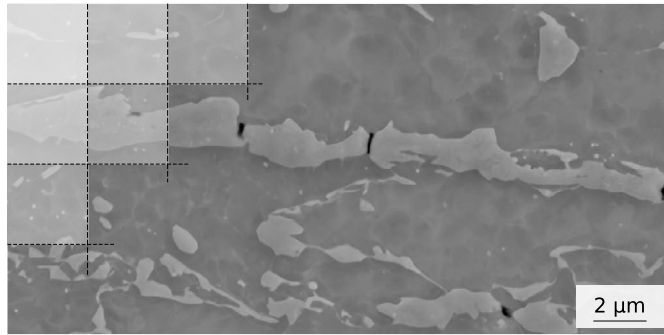


Fig. 7. Example of super-resolution patches, stitched together to a larger micrograph.

Table 3

Evaluation metrics from the dual-phase steel dataset using a window size of 4096×4096 pixels.

Metric	Bi-Cubic	Lanczos	TTSR
SSIM	0.559 ± 0.005	0.546 ± 0.005	0.574 ± 0.005
PSNR	$22.1 \text{ dB} \pm 0.1 \text{ dB}$	$21.8 \text{ dB} \pm 0.1 \text{ dB}$	$22.2 \text{ dB} \pm 0.1 \text{ dB}$
ESSIM	0.353 ± 0.005	0.340 ± 0.005	0.400 ± 0.006

3.2. Evaluation of network performance on case-hardening steel

To further investigate the transferability of the developed methodology, we applied the network trained on dual-phase steel data directly to the case-hardening steel dataset before performing additional training using a small dataset of 1600 image pairs and assessed the improvement.

3.2.1. Baseline

To assess the performance of the network, we again established a baseline for comparison. An example of the typical microstructure for this material is shown in Fig. 8. Here, we illustrated the performance of nearest, bi-cubic, and Lanczos interpolation alongside the high-resolution ground truth images. The cementite lamellae were not visible in the nearest-neighbour interpolation and are barely visible both for bi-cubic and Lanczos interpolation. The similarity metrics for Lanczos and bi-cubic interpolation are summarised in Table 4. As bi-cubic

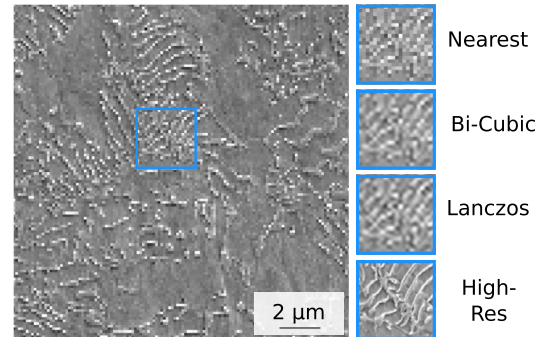


Fig. 8. Low-resolution input and comparison of nearest-neighbour, bi-cubic, and Lanczos interpolation with the high-resolution ground truth of the area highlighted in blue on a 16MnCrS5 case-hardening steel scanning electron micrograph.

interpolation performed better than Lanczos interpolation, we chose this for the comparisons with the machine-learning based super-resolution method.

3.2.2. Direct application to case-hardening steel micrographs

To test whether additional training is necessary when changing material systems or a substitution of reference images is sufficient, we used the network weights trained on the dual-phase steel, supplied high-resolution reference images from the 16MnCrS5 dataset and evaluated the performance of the network. The direct application of the network trained on the dual-phase steel dataset produced overly bright images, with a mean pixel value of 202, compared to a mean pixel value of 133 in the ground truth dataset. In order to evaluate whether this is related to the brightness and/or contrast differences during image acquisition, we re-scaled the pixel values of the images after resolution-enhancement to the high-resolution ground truth distribution. However, this resulted in images that were still overly bright and showed additional colour artefacts as can be seen in Fig. 9 (c). Furthermore, the features at the borders of the predictions did not align leading to stitching artefacts, which can be seen as straight lines in both horizontal and vertical direction in the enhanced image, highlighted with blue arrows in Fig. 9 (c).

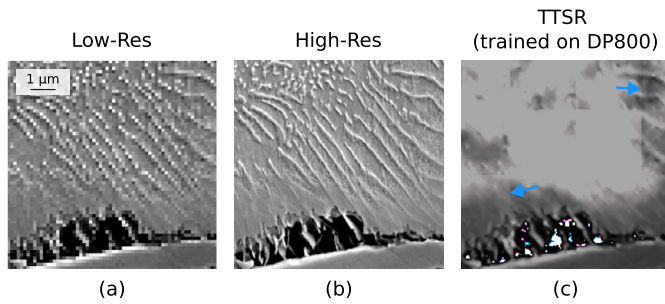


Fig. 9. Resolution-enhancement network trained on dual-phase steel dataset applied to 16MnCrS5 case-hardening steel micrographs without further network training, (a) low-resolution input, (b) high-resolution ground truth, (c) output of the texture transformer. Stitching artefacts are indicated with blue arrows.

Table 4

Similarity metrics for bi-cubic, Lanczos interpolation and super-resolution on the aligned 16MnCrS5 dataset.

Metric	Bi-Cubic	Lanczos	TTSR (fine-tuned)
SSIM	0.476 ± 0.003	0.467 ± 0.003	0.492 ± 0.004
PSNR	$18.4 \text{ dB} \pm 0.1 \text{ dB}$	$18.1 \text{ dB} \pm 0.1 \text{ dB}$	$19.1 \text{ dB} \pm 0.1 \text{ dB}$
ESSIM	0.314 ± 0.004	0.309 ± 0.004	0.488 ± 0.006

3.2.3. Fine-tuning with data from case-hardening steel

Fig. 9 suggests that the network trained on dual-phase steel cannot be applied directly to a new material but fine-tuning is necessary when changing material systems. We fine-tuned with a small dataset consisting of 1600 low-resolution, high-resolution, and reference images, corresponding to approximately 20 min of recording time.

The resulting similarity metrics are summarised in Table 4. The fine-tuned network improved both in terms of SSIM, PSNR, and ESSIM over Lanczos and bi-cubic interpolation.

A visual comparison between different microstructural elements from the low-resolution input, high-resolution ground truth, the output of the TTSR network, and the output of bi-cubic interpolation is shown in Fig. 10. Fig. 10 (a) and (b) depict two examples of pearlitic regions. The network produced clearer cementite lamellae that closely resemble the high-resolution ground truth. Fig. 10 (b) illustrates how challenging the enhancement of the fine lamellae was. Here, the resolution-enhancement algorithm showed a continuous lamella whereas there is a small gap in the high-resolution image. However, we could see from the low-resolution input, that the reconstruction of such intricate details is challenging. In Fig. 10 (c) and (d) different MnS inclusions are displayed. The network performed seemingly better when reconstructing high-resolution information in these MnS inclusions compared to the fine pearlitic structures.

4. Discussion

In this work we have successfully implemented and trained a deep learning reference-based super-resolution algorithm for scanning electron microscopy using pre-trained model architectures. We recorded three datasets of scanning electron micrographs on two materials with visually different microstructures, with a target resolution commonly used for the damage analysis in each respective material and the input resolution of the algorithm at a quarter of that value. Here, we discuss the performance and limitations of the developed methodology, difficulties with similarity metrics for scanning electron micrographs of metallic microstructures, and present a novel imaging workflow based on the texture transformer network.

4.1. Metrics for resolution enhancement

We quantitatively assess the improvement of the enhanced image compared to the high-resolution ground truth with multiple metrics: SSIM, PSNR, and ESSIM. While SSIM and PSNR show modest improvements when using the machine-learning based approach compared to the conventional interpolation methods, the edge-based structural similarity index demonstrates more substantial enhancements. This aligns better with our visual assessment where machine-learning enhanced images appear considerably closer to the high-resolution ground truth than interpolated ones. The discrepancy between metrics can be attributed to the different approaches followed by interpolation and resolution-enhancement: In terms of traditional metrics, interpolation is a good conservative approximation of the high-resolution images, where coinciding pixels have the same value. When using an interpolation method, these original pixels form the “scaffolding” that is then filled by the interpolation algorithm using a smooth function. The machine-learning based method, on the other hand, creates a new image and, therefore, does not simply fill in pixels in the space between pixels from the original low-resolution input. In addition, the resolution-enhancement algorithm is trained with a more complex mixture of metrics in addition to the reconstruction loss that compares the pixel-wise difference between images. Therefore, the images follow a less conservative approach after resolution-enhancement, resulting in sharper features in contrast to the smooth interpolation. As these sharp features correspond to fast changes in pixel values, small deviations in position have a strong negative impact on the conventional similarity metrics. The ESSIM metric, which specifically emphasizes edge preservation, better captures these improvements in structural details and boundary definition that are critical for accurate microstructural characterisation. Additionally, similarity metrics are not designed to take noise into account that is inherent to electron micrographs. However, as this noise is, essentially, a physical artefact from the image acquisition process, comparing two micrographs is not entirely suitable to assess super resolution performance. In particular, because small shifts in the noisy pixels do not change the underlying image information in any way but may influence the numerical value of the metrics. While extensions of both the SSIM and PSNR exist that were designed to be less sensitive to noise, e.g. PSNR-HVS, SSIM-HVS, PSNR-HVS-M, SSIM-HVS-M [36,37], these extensions are designed after the perception of the human visual system and introduce additional biases in their calculation. The ESSIM metric, with its focus on edge preservation, provides a more relevant evaluation for our microstructural analysis without introducing such perceptual biases.

4.2. Feature size constraints and physical validity in deep learning-based SEM image enhancement

The application of deep-learning based super-resolution to SEM imaging reveals important considerations regarding the relationship between physical feature sizes and reconstruction quality. Our observations across dual-phase and case-hardening steel samples demonstrate that special attention must be paid when the dimensions of microstructural features approach the resolution limit of the low-resolution input images. This becomes particularly evident in the reconstruction of pearlitic structures in the 16MnCrS5 case-hardening steel, where cementite lamellae dimensions are close to the lower resolution threshold. In Fig. 10 (b), for example, the network predicted a continuous lamella structure where the ground truth revealed two distinct lamellae meeting at a point. This discrepancy highlights a fundamental challenge: when feature sizes approach the resolution limit, the network may generate plausible but physically incorrect reconstructions in the absence of constraining factors. Interestingly, this limitation appears to be mitigated in cases where physical constraints are inherently represented in the training dataset. For example, the network successfully reconstructed martensite cracks, Fig. 6 (a), despite their similar physical dimensions. This success can be attributed to the consistent morphological

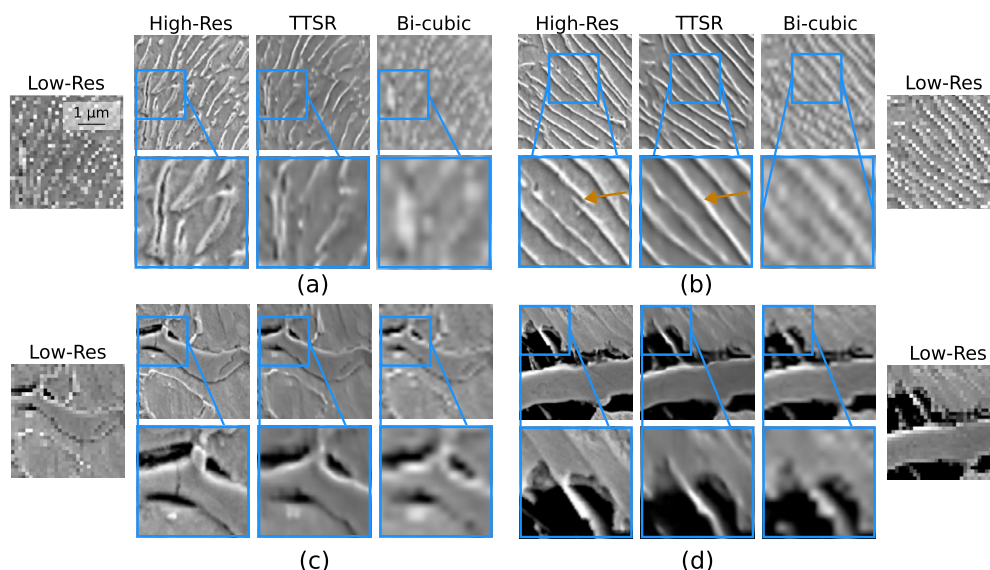


Fig. 10. Different points of interest of the microstructure of the case-hardening steel shown for the low-resolution input, the high-resolution ground truth, the output of the network, and for comparison the bi-cubic interpolation results. Each row depicts a larger area in high-resolution and one zoomed in region.

patterns of cracks present in the training data, effectively providing the network with implicit physical constraints. The contrast between these cases demonstrates that the network's ability to make accurate predictions depends not only on feature size but also on whether the physical constraints are adequately expressed through examples in the training dataset. These observations underscore the importance of careful validation when applying deep learning-based super resolution to microstructural analysis, particularly when examining features near the resolution limit.

4.3. Cross-material adaptability and transfer learning

The direct application of the texture-transformer trained on the dual-phase steel dataset resulted in overly bright images, even after re-scaling of brightness and contrast. Furthermore, stitching of the predicted patches to a larger image resulted in stitching artifacts present at the borders of the prediction. We attribute this to the substantially different microstructure of the 16MnCrS5 case-hardening steel. Unifying the imaging conditions between different materials could mitigate this issue to some extent.

Several factors likely influenced this transferability challenge, with microstructural differences being the most significant. The fundamental disparity between dual-phase steel (ferrite-martensite structure) and 16MnCrS5 (ferrite, pearlitic regions and MnS inclusions) creates distinctly different feature morphologies, size distributions, and boundary characteristics. As shown in Fig. 10, feature complexity significantly impacts transferability—the model adapted more readily to MnS inclusions than to the intricate pearlitic lamellae structures. A multitude of factors beyond the microstructure itself may impact transferability, but would require specific and in some case correlative new datasets, e.g. use of different electron detectors (we exclusively used SE imaging in this work), sample preparation (our microstructures were polished and etched for phase contrast), or simply imaging parameters such as acceleration voltage. While we demonstrate successful transfer to reasonably similar images taken from very different microstructures, limitations may arise in future applications across diverse materials and microscopes. In any case, we assume that the collection of fine-tuning data that adequately represents all key microstructural features of the target material will be beneficial or even essential.

Using a smaller dataset consisting of 2000 additional images, corresponding to 20 min of image acquisition time, for further training of the neural network improved the performance of the network substantially,

both in terms of the SSIM, PSNR, and ESSIM metrics, as well as visual similarity.

This successful adaptation through fine-tuning suggests a promising pathway for broader application of machine learning-based resolution enhancement in electron microscopy. Recent advances in foundation models, such as the Segment Anything Model [38], have demonstrated the power of pre-trained models that can be rapidly adapted to new tasks. Drawing inspiration from these developments, we envision the creation of a foundation SEM enhancement model trained on a comprehensive dataset spanning various materials and imaging conditions. Such a model could capture universal features of electron microscopy images. Our results suggest that this foundation model could then be efficiently fine-tuned to new materials and imaging conditions with minimal additional data collection, potentially requiring even fewer than the 2000 images used in this study. This approach could significantly streamline the adoption of super-resolution techniques in electron microscopy while maintaining high enhancement quality across different material systems.

Beyond material-specific adaptability, we also investigated the effectiveness of our approach when applied to images acquired with accelerated scanning parameters. A supplementary experiment on the dual-phase steel with reduced dwell time (10 μ s versus 100 μ s) demonstrated that the texture transformer can effectively enhance noisier, rapidly acquired micrographs while still outperforming traditional interpolation methods (see Table S1 in the supplementary material section A). This suggests our approach can be combined with parameter optimisation strategies, offering independent yet complementary pathways to reduce overall acquisition time in SEM imaging.

4.4. Experimental acceleration through intelligent scanning strategies

The application of deep learning-based resolution enhancement transforms microscopy workflows by reducing image acquisition time by a factor of 16. This acceleration converts what would traditionally be a 9 h scan of a 1 mm² area at 32.5 nm/px resolution into a 30 min procedure. While this speed-up alone enables previously infeasible experiments, such as complete cross-section analysis of macroscopic tensile specimens or multi-step in-situ deformation studies within standard 8 h instrument sessions, its true potential lies in combining rapid acquisition with intelligent scanning strategies.

Drawing parallels from face recognition research [39], where enhanced low-resolution images can improve feature detection despite

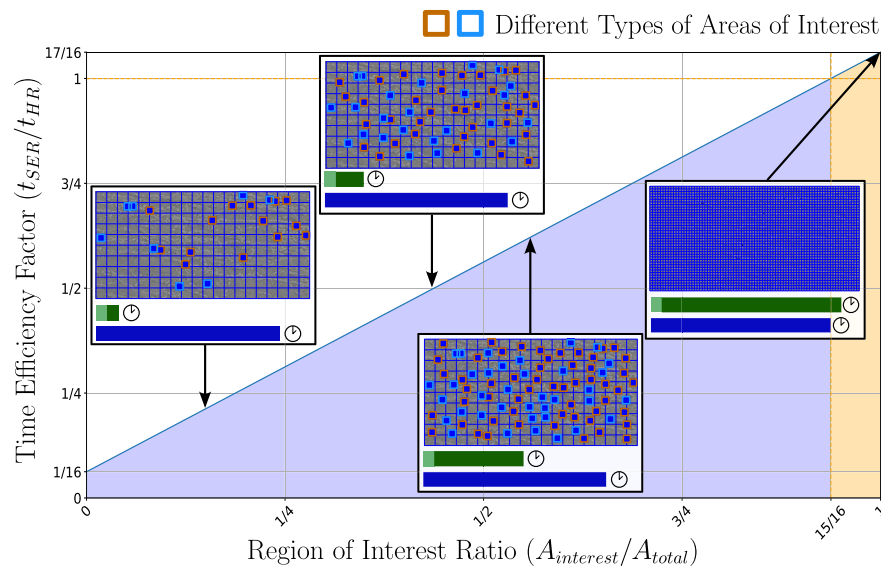


Fig. 11. Depiction of the expected relative imaging time compared to conventional high-resolution scanning (y-axis) against the fraction of the area of interest over the total area to be imaged (x-axis). The ratio between imaging times using super-resolution and imaging with full high-resolution depends linearly on the area that needs to be rescanned and the total area. Several micrographs depicting $A_{\text{interest}}/A_{\text{total}}$ scenarios are given as well. In regions of the graph highlighted in blue time can be saved by applying a scan-enhance-rescan strategy.

not being perfectly accurate, we propose a scan-enhance-rescan (SER) strategy. In this approach, the resolution-enhanced micrographs serve as intelligent guides for identifying regions of interest, even if some fine details may not be perfectly reconstructed. The enhanced images, having re-introduced features such as clear phase boundaries and more distinct voids, enable more precise targeting of subsequent high-resolution scans. This approach can be further refined by quantifying model uncertainty in the enhanced images, prioritizing areas of high uncertainty for rescanning to ensure physical realism of critical features.

For a resolution ratio of 4 between high- and low-resolution scans in each dimension, the time efficiency of this strategy can be expressed as:

$$t_{\text{SER}}/t_{\text{HR}} = 1/16 + A_{\text{interest}}/A_{\text{total}} \quad (10)$$

where t_{SER} is the combined time for a scan-enhance-rescan strategy, i.e. low-resolution scanning and high-resolution re-scanning of regions of interest, t_{HR} is the time for complete high-resolution scanning, A_{interest} is the combined area of regions of interest, and A_{total} is the total scan area. This approach remains time-efficient up to a region of interest to total area ratio of 15/16, making it practical even when many regions require detailed examination. A visual depiction of these estimates can be found in Fig. 11. The method is particularly valuable for in-situ experiments, where high-resolution images from previous steps can serve as reference data for the model, improving subsequent predictions. This creates a positive feedback loop where each experimental stage benefits from prior high-quality data while maintaining rapid acquisition speeds.

5. Conclusion

In this work, we developed a novel resolution-enhancement method for improving scanning electron micrographs that generates much more realistic images compared to conventional interpolation approaches. The method, based on a texture-transformer deep-learning algorithm, was able to enhance low-resolution micrographs with high fidelity for the subsequent analysis that requires high-resolution images.

By reducing the recording resolution we achieved a 16-fold decrease in acquisition time while maintaining image quality suitable for microstructural analysis.

We further show that the trained network is capable of adapting to a visually different microstructure through fine-tuning with a dataset associated with an additional recording time of 20 min.

Furthermore, we propose a scan-enhance-rescan strategy in cases where accurate high resolution images of rare features are essential for subsequent analysis. In this case, low resolution scanning coupled with resolution enhancement can serve as a starting point for the localisation of points of interest, effectively avoiding lengthy high-resolution imaging on sample area not of interest. Following this approach, large-area in-situ and slice-and-view experiments can be performed in a fraction of the previously required time. In this way, a larger range of compositions, processing routes, strains, or simply observation of more representative areas, to name just a few examples, become accessible by electron microscopy.

CRedit authorship contribution statement

T. Reclik: Writing – original draft, Software, Methodology, Investigation, Data curation. **S. Medghalchi:** Writing – review & editing, Methodology. **P. Schumacher:** Methodology, Investigation. **M.A. Woltenweber:** Writing – review & editing, Data curation. **T. Al-Samman:** Writing – review & editing, Supervision. **S. Korte-Kerzel:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization. **U. Kerzel:** Writing – review & editing, Supervision, Project administration, Funding acquisition, Conceptualization.

Data and code availability

The datasets used in this work can be found on Zenodo <https://doi.org/10.5281/zenodo.15237354>.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used Claude (claude.ai) in order to improve language and readability. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors express their gratitude to the Deutsche Forschungsgemeinschaft (DFG) for financial support in context of the Collaborative Research Centre CRC/Transregio 188/2 “Damage Controlled Forming Processes”, project T02 and B02, project no. 278868966. Simulations were performed with computing resources granted by WestAI under project rwth1596.

Appendix A. Supplementary material

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.matdes.2025.113955>.

References

- [1] T. Lei, D. Wang, X. Yu, S. Ma, W. Zhao, C. Cui, J. Meng, S. Tao, D. Guan, Global iron and steel plant CO₂ emissions and carbon-neutrality pathways, *Nature* 622 (Oct. 2023) 514–520, Publisher: Nature Publishing Group.
- [2] International Energy Agency, CO₂ Emissions in 2023, report, IEA, Feb. 2024.
- [3] International Energy Agency, Iron and Steel Technology Roadmap: Towards More Sustainable Steelmaking, report, IEA, 2020.
- [4] J. Kang, Microscopic Strain Localization and Damage in Multi-phase Alloys, Thesis, Massachusetts Institute of Technology, Sept. 2022, Accepted: 2023-01-19T18:39:47Z.
- [5] R. Gitschel, O. Hering, A. Schulze, A. Erman Tekkaya, Controlling damage evolution in geometrically identical cold forged parts by counterpressure, *J. Manuf. Sci. Eng.* 145 (Dec. 2022).
- [6] A.E. Tekkaya, J.M. Allwood, P.F. Barianti, S. Bruschi, J. Cao, S. Gramlich, P. Groche, G. Hirt, T. Ishikawa, C. Löhbe, J. Lueg-Althoff, M. Merklein, W.Z. Misiolek, M. Pietrzyk, R. Shivpuri, J. Yanagimoto, Metal forming beyond shaping: predicting and setting product properties, *CIRP Ann.* 64 (Jan. 2015) 629–653.
- [7] C. Kusche, T. Reclik, M. Freund, T. Al-Samman, U. Kerzel, S. Korte-Kerzel, Large-area, high-resolution characterisation and classification of damage mechanisms in dual-phase steel using deep learning, *PLoS ONE* 14 (May 2019) e0216493, Publisher: Public Library of Science.
- [8] C. Tian, C.F. Kusche, A. Medina, S. Lee, M.A. Wollenweber, R. Pippan, S. Korte-Kerzel, C. Kirchlechner, Understanding the damage initiation and growth mechanisms of two DP800 dual phase grades, *Mater. Des.* 238 (Feb. 2024) 112630.
- [9] M.A. Wollenweber, C.F. Kusche, T. Al-Samman, S. Korte-Kerzel, On the automated characterisation of inclusion-induced damage in 16MnCrS5 case-hardening steel, *Adv. Ind. Manuf. Eng.* 7 (Nov. 2023) 100123.
- [10] S. Medghalchi, J. Kortmann, S.-H. Lee, E. Karimi, U. Kerzel, S. Korte-Kerzel, Automated segmentation of large image datasets using artificial intelligence for microstructure characterisation and damage analysis, *Mater. Des.* 243 (July 2024) 113031.
- [11] S. Medghalchi, C.F. Kusche, E. Karimi, U. Kerzel, S. Korte-Kerzel, Damage analysis in dual-phase steel using deep learning: transfer from uniaxial to biaxial straining conditions by image data augmentation, *JOM* 72 (Dec. 2020) 4420–4430.
- [12] F. Yang, H. Yang, J. Fu, H. Lu, B. Guo, Learning Texture Transformer Network for Image Super-Resolution, *IEEE Computer Society*, June 2020, pp. 5790–5799.
- [13] Z. Zhang, Z. Wang, Z. Lin, H. Qi, Image Super-Resolution by Neural Texture Transfer, *IEEE Computer Society*, June 2019, pp. 7974–7983.
- [14] H. Chen, Y. Zhang, M.K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, G. Wang, Low-dose CT with a residual encoder-decoder convolutional neural network, *IEEE Trans. Med. Imaging* 36 (Dec. 2017) 2524–2535.
- [15] Y. Chen, Y. Xie, Z. Zhou, F. Shi, A.G. Christodoulou, D. Li, Brain MRI super resolution using 3D deep densely connected neural networks, in: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Apr. 2018, pp. 739–742, arXiv:1801.02728 [cs, eess].
- [16] J. Xu, E. Gong, J. Pauly, G. Zaharchuk, 200x low-dose PET reconstruction using deep learning, arXiv:1712.04119 [cs], Dec. 2017.
- [17] K. de Haan, Z.S. Ballard, Y. Rivenson, Y. Wu, A. Ozcan, Resolution enhancement in scanning electron microscopy using deep learning, *Sci. Rep.* 9 (Aug. 2019) 12050, Number: 1, Publisher: Nature Publishing Group.
- [18] Z. Lu, J. Li, H. Liu, C. Huang, L. Zhang, T. Zeng, Transformer for Single Image Super-Resolution, *IEEE Computer Society*, June 2022, pp. 456–465.
- [19] C. Dong, C.C. Loy, K. He, X. Tang, Learning a deep convolutional network for image super-resolution, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), *Computer Vision – ECCV 2014*, Cham, Springer International Publishing, 2014, pp. 184–199.
- [20] C.C. Tasan, J.P.M. Hoefnagels, M. Diehl, D. Yan, F. Roters, D. Raabe, Strain localization and damage in dual phase steels investigated by coupled in-situ deformation experiments and crystal plasticity simulations, *Int. J. Plast.* 63 (Dec. 2014) 198–210.
- [21] S. Medghalchi, E. Karimi, S.-H. Lee, B. Berkels, U. Kerzel, S. Korte-Kerzel, Three-dimensional characterisation of deformation-induced damage in dual phase steel using deep learning, *Mater. Des.* 232 (Aug. 2023) 112108.
- [22] K. Isik, G. Gerstein, T. Clausmeyer, F. Nürnberger, A.E. Tekkaya, H.J. Maier, Evaluation of void nucleation and development during plastic deformation of dual-phase steel DP600, *Steel Res. Int.* 87 (12) (2016) 1583–1591, eprint, <https://onlinelibrary.wiley.com/doi/pdf/10.1002/srin.201500483>.
- [23] H.R.V. Joze, I. Zharkov, K. Powell, C. Ringler, L. Liang, A. Roulston, M. Lutz, V. Pradeep, ImagePairs: Realistic Super Resolution Dataset via Beam Splitter Camera Rig, *IEEE Computer Society*, June 2020, pp. 2190–2200.
- [24] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 3rd International Conference on Learning Representations (ICLR 2015), Computational and Biological Learning Society, 2015.
- [25] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, ImageNet: a large-scale hierarchical image database, in: 2009 IEEE Conference on Computer Vision and Pattern Recognition, June 2009, pp. 248–255, ISSN: 1063-6919.
- [26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, L. Kaiser, I. Polosukhin, Attention is all you need, in: *Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS’17*, Red Hook, NY, USA, Curran Associates Inc., Dec. 2017, pp. 6000–6010.
- [27] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, in: B. Leibe, J. Matas, N. Sebe, M. Welling (Eds.), *Computer Vision – ECCV 2016*, Springer International Publishing, Cham, 2016, pp. 694–711.
- [28] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, *Adv. Neural Inf. Process. Syst.* 3 (June 2014).
- [29] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein generative adversarial networks, in: *Proceedings of the 34th International Conference on Machine Learning, PMLR, July 2017*, pp. 214–223, ISSN: 2640-3498.
- [30] L. Metz, B. Poole, D. Pfau, J. Sohl-Dickstein, Unrolled generative adversarial networks, arXiv:1611.02163 [cs, stat], May 2017.
- [31] L.N. Smith, Cyclical learning rates for training neural networks, in: 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 2017, pp. 464–472.
- [32] D.P. Kingma, J. Ba, Adam: a method for stochastic optimization, *CoRR*, arXiv:1412.6980 [cs.LG], Dec. 2014.
- [33] C.E. Duchon, Lanczos filtering in one and two dimensions, *J. Appl. Meteorol. Climatol.* 18 (Aug. 1979) 1016–1022, Publisher: American Meteorological Society Section.
- [34] Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (Apr. 2004) 600–612, Conference Name: IEEE Transactions on Image Processing.
- [35] G.-H. Chen, C.-L. Yang, L.-M. Po, S.-L. Xie, Edge-based structural similarity for image quality assessment, in: 2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, vol. 2, May 2006, pp. II-933–II-936, ISSN: 2379-190X.
- [36] K. Egiazarian, J. Astola, V. Lukin, F. Battisti, M. Carli, New full-reference quality metrics based on HVS, in: *Proceedings of the Second International Workshop on Video Processing and Quality Metrics*, Jan. 2006.
- [37] N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin, On between-coefficient contrast masking of DCT basis functions, in: *Proc of the 3rd Int Workshop on Video Processing and Quality Metrics for Consumer Electronics*, Jan. 2007.
- [38] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A.C. Berg, W.-Y. Lo, P. Dollár, R. Girshick, Segment anything, in: 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 2023, pp. 3992–4003, ISSN: 2380-7504.
- [39] M.P. Autee, M.S. Mehta, M.S. Desai, V. Sawant, A. Nagare, A review of various approaches to face hallucination, *Proc. Comput. Sci.* 45 (Jan. 2015) 361–369.