



Application of artificial intelligence to determine the phase fraction of welded duplex steels using neural networks with partial use of augmented data

Leon Kaufhold¹, Julia Wichmann¹, Galina Polte², and Thomas Schroeter¹

¹Günter Köhler Institute for Joining Technology and Materials Testing,
Ernst-Ruska-Ring 3, 07745 Jena, Germany

²Faculty of Mechanical Engineering, Group of Quality Assurance and Industrial Image Processing,
Ilmenau University of Technology, Gustav-Kirchhoff-Platz 2, 98693 Ilmenau, Germany

Correspondence: Thomas Schroeter (foko@ifw-jena.de)

Received: 11 November 2025 – Revised: 15 February 2026 – Accepted: 21 February 2026 – Published: 13 July 2026

Abstract. This paper addresses the application of artificial intelligence in metallography. The aim is to determine the phase fractions of ferrite and austenite in the weld zone of duplex stainless steel based on the evaluation of metallographic microscopic micrographs. The idea arose from the motivation to technically replicate the functionality of the biological eye–brain system, thereby combining the properties of the highly effective human vision (or its simulation) with the significantly faster technical image-processing system. Since the distribution of pixel amplitudes in the metallographic images is irregular, the use of artificial intelligence is a suitable alternative to algorithm-based or gray-value statistical methods. Butt-welded sheet metal strips, which were cut in the weld zone, were used as sample material. The cut surfaces were polished and color-etched. Microscopic images were taken of the prepared samples. In the subsequent image analysis, both simple perceptron variants and deep convolutional neural networks (CNNs) were used. These were designed using appropriate tools and (subsequently) implemented programmatically. To obtain a sufficient amount of training data, the use of synthetic data derived from original data was investigated and implemented for training the networks in addition to using polished samples from real specimens. Training and test runs were conducted with various datasets and network variants, and performance parameters were determined. The results obtained met expectations regarding both the use of synthetic data and the classification error rates of 2 %–3 %. Significant performance differences were observed in processing speed. Here, the convolutional neural networks performed considerably better than the perceptrons.

1 Introduction

In the context of climate and environmental aspects, the use of innovative materials, such as high-strength lightweight steels, is highly significant. This includes their production (“green steel”), the efficient use of resources, and the design of their material properties. State-of-the-art process control methods are now available for the production of these materials. These methods allow for the precise control and scaling of microstructures and their formation during metal production, enabling the precise adjustment of desired material properties such as strength, elasticity, yield strength, and

phase structure. Quality control is carried out by means of, among other methods, the metallographic analysis of micrographs of the steels. In addition to the steel itself, the microstructure properties in the weld zone also play a crucial role in welded joints.

Digital image processing (DIP) combined with artificial intelligence (AI) is increasingly being used to analyze material structures. Artificial neural networks are a key AI application. They can solve complex problems that are difficult to handle with rule-based algorithms. The first applications of neural networks emerged in the late 1980s with the availability of suitable hardware platforms offering high stor-

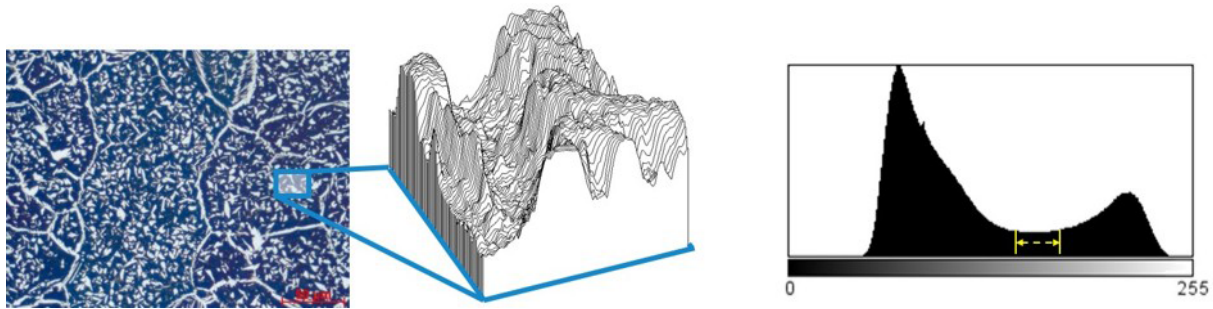


Figure 1. Microscopic micrograph of duplex stainless steel in the weld zone at 500 $\times$ magnification and a section of the height profile of the gray-scale pixel amplitudes (left panel). Histogram of the gray-scale pixel amplitudes of the micrograph (right panel).

age capacity and computing power. A further breakthrough came around 2012 with the development of deep neural networks with many layers and convolutional networks, which perform mathematical convolution operations between the layers. This paper deals with the determination of the phase fractions of austenite and ferrite in the weld zone of duplex steel. Investigations using both perceptrons and folding deep-structured neuronal networks are described.

Figure 1 shows a metallographic image of the weld zone of a duplex stainless steel and, on the left, a section of the corresponding gray-scale intensity profile. Ferrite areas are shown in blue, and austenite areas are shown in white. The histogram of the gray-scale pixel amplitudes is shown on the right. The aim is to precisely determine the phase fractions of ferrite and austenite in the weld zone by evaluating metallographic images (Fig. 1, left panel). As can be seen in the height profile in Fig. 1, the pixel amplitudes exhibit irregular structures. Therefore, algorithmic methods are not suitable for segmenting the phase fractions. The histogram of the pixel amplitudes (Fig. 1, right panel) shows only a weakly pronounced minimum. Threshold methods or their modifications (Otsu, 1979) are suitable for low-accuracy requirements; for higher-accuracy requirements, they are rather unsuitable. Due to the described image properties, it was proposed to solve the problem using artificial intelligence.

2 Sample material, sample preparation, and objects for training and evaluation

The samples to be analyzed were tungsten inert gas (TIG)-welded joints of duplex steel (Ulrich et al., 2019). Duplex steel contains austenite, as well as ferrite. During welding, the ratio of both parts in the weld zones change (Folkhard, 1983). Thermal vibrations during welding can shift the ferrite–austenite ratio. Standards typically specify 30 %–65 % ferrite in the heat-affected zone (HAZ), with extended tolerances of up to 70 % for weld areas. A ferrite content exceeding 70 % can result in a completely ferritic microstructure, significantly impairing toughness and corrosion resistance. Therefore, determining the phase ratios in the weld

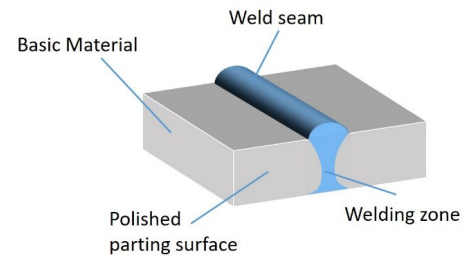


Figure 2. Specimen with joint and weld zone (schematic representation).

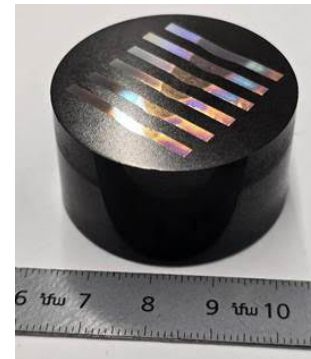


Figure 3. Embedding of cross-section samples, ground, polished, and color etched.

is of great importance. For this investigation, strips of 1.4462 (UNS 2205) were butt-welded at the end faces. Subsequently, 15 mm $\times$ 8 mm specimens were cut out to expose the cross-section of the weld zone (see Fig. 2 for a schematic representation). The prepared samples were embedded in epoxy resin under heat (180 $^{\circ}$ C) and pressure (250 bar) to create a 40 mm $\varnothing$ embedding with good etch retention, containing up to 6 cross-sectional pieces (see Fig. 3).

The samples were wet-ground with 180-grit Silicon carbide (SiC) abrasive grains sandpaper and then fine-ground to 1200 grit. Polishing was performed with diamond paste down to a particle size of 3 μ m. To achieve optimal contrast be-

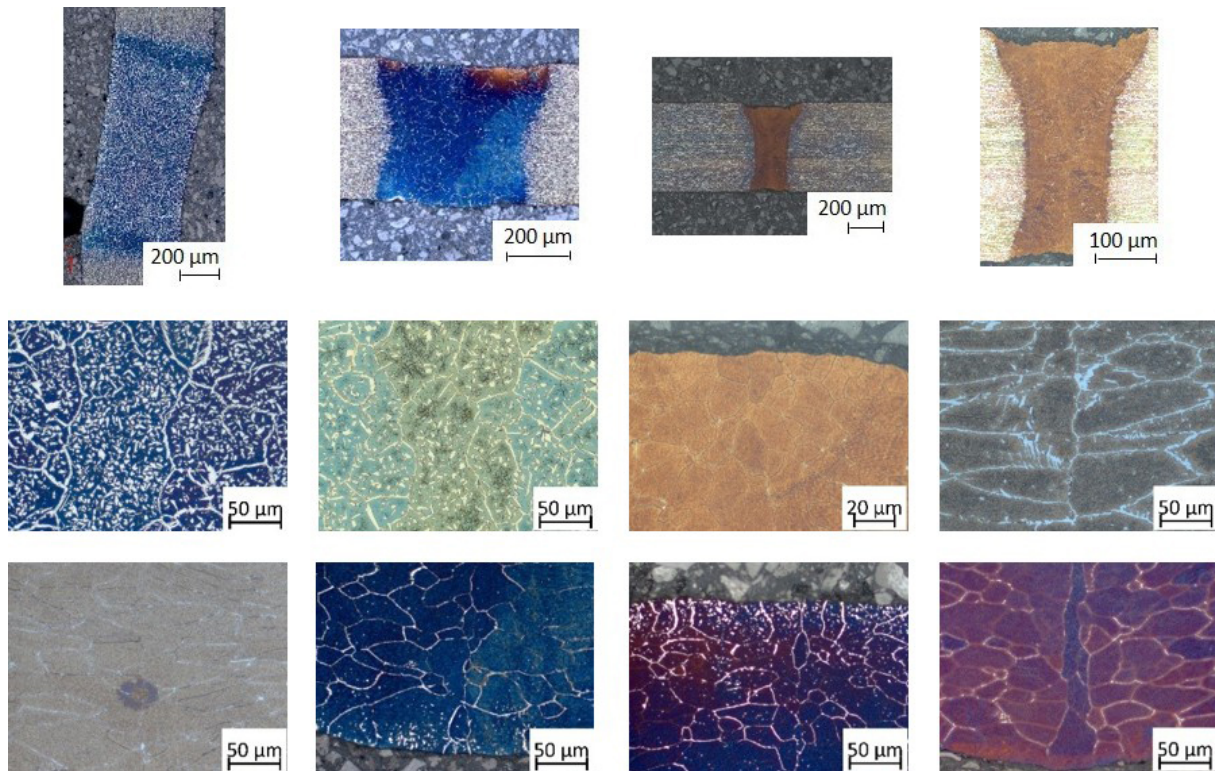


Figure 4. Color-etched micrographs of TIG-welded duplex stainless steel specimens. First row: overview images of the entire weld zone. Second and third row: 500 $\times$ magnification sections.

tween the austenite and ferrite components, the samples were color-etched with Beraha-II etchant (Beraha, 1968). Light microscopy images of the samples prepared in this way were acquired using ZEISS Axioskop 7 and ZEISS Axiocam 705 color and digitized as RGB images with a color depth of 32 bits (approximately 16.7 million color combinations) and an image size of 2584×1936 pixels (5 MP) in JPEG format.

Figure 4 shows examples of color-etched photomicrographs and microscopically magnified sections.

Especially underrepresented classes are often a problem. For example, during the production of test samples, some defects to be classified occur only rarely and cannot be deliberately induced. In this case, however, the problem is the insufficient number of test objects (micrographs) available at all. One possible solution is to use multiple micrographs. However, the production of color-etched micrographs is very complex. At reasonable costs, only single-digit to low-double-digit quantities of test samples could be produced. The preparation of samples is rather complex due to the fact that the same microstructure can produce different images in terms of color and structure (different types; see Fig. 4). In the work described here, only images of two types were initially used.

3 Image augmentation

Image augmentation is one approach to solving the problem of underrepresented classes or too few test objects for the training process. Two of the most common methods for expanding datasets can be distinguished: data transformation and data synthesis (Cogswell et al., 2015; Wunsch et al., 2023). Data transformation involves applying minor modifications to existing original data in order to increase its variability without significantly changing the fundamental information structure. In contrast, data synthesis refers to the artificial generation of data that are similar in terms of their characteristics to the original dataset but not identical to it. The following methods were considered for metallographic applications in our studies:

- classical methods (transformation, non-destructive) or simple mathematical methods, e.g., rotation by various angles, reflections, etc., and
- GAN methods using deep neural networks (Generative Adversarial Networks, GANs) (Goodfellow et al., 2014).

The GAN analyses were conducted using the official PyTorch implementation of StyleGAN3 (Karras et al., 2019) and a pre-trained LHQ 256 model (Karras et al., 2021) on a

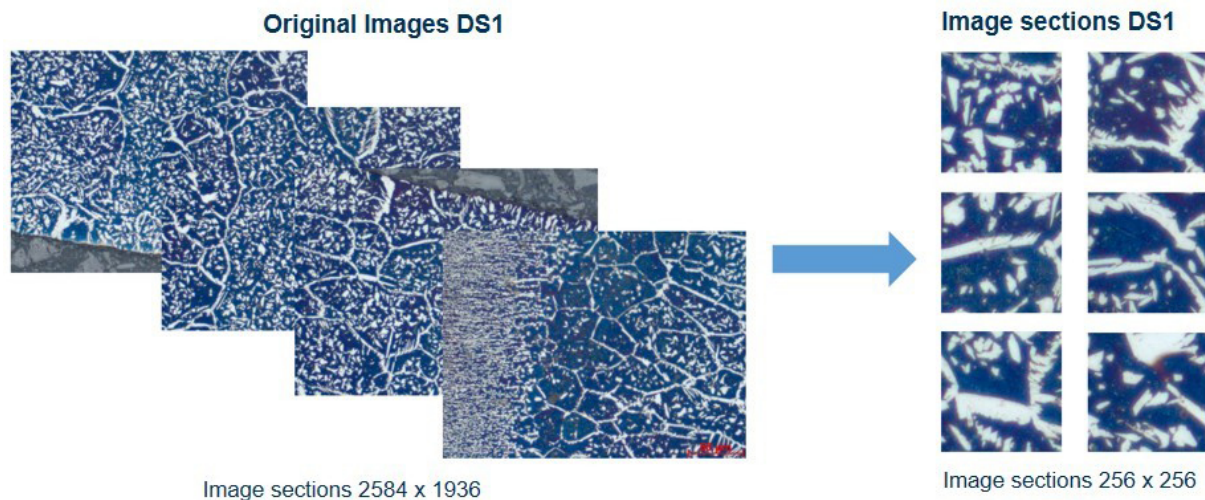


Figure 5. Images and image sections, dataset 1.

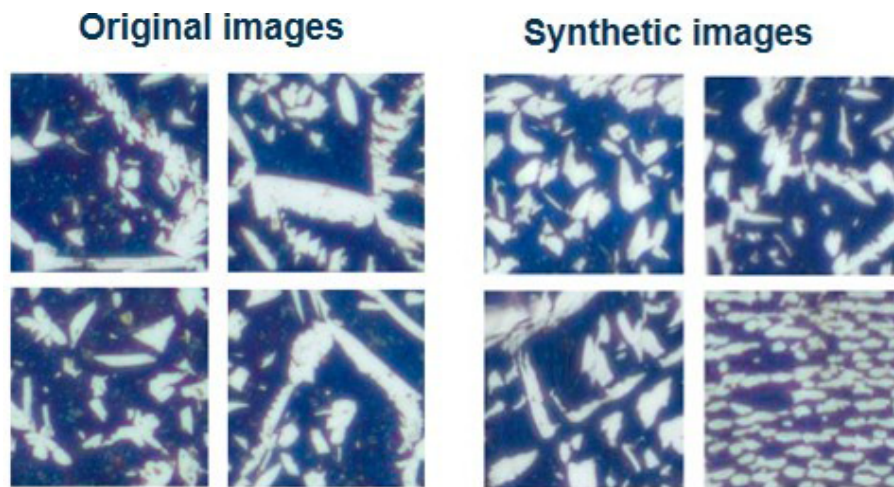


Figure 6. Results of image augmentation, dataset 1. GAN used: StyleGAN3 (Karras et al., 2021). Image size: 256×256 . Calculation on a computing cluster at TU Ilmenau, official PyTorch implementation pre-trained model LHQ-256 (Karras et al., 2021).

computing cluster at the TU Ilmenau. A total of three datasets of metallographic images were analyzed. The results of two datasets are presented below. Figure 5 shows the original images of a dataset used for augmentation, along with corresponding image sections.

Sections of 256×256 pixels were cut from the original images and fed into the network for augmentation. Figure 6 shows selected augmentation results in the form of 256×256 synthetic images. A total of 268 image sections were used to train StyleGAN3.

Up to a training progress of $\text{kim} = 1040$, consistent and qualitatively satisfactory results were achieved. One Kim corresponds to 1000 images processed by the model during training (see Fig. 6). Beyond this value, however, the training no longer converged, resulting in images with completely

different compositions (see Fig. 7). The results in the form of the generated synthetic images were used in the following work, along with the images of real specimens, for training and testing the neural networks for the metallographic tasks. Figures 8 and 9 show the original data and corresponding image excerpts from the second dataset. A purely visual inspection revealed respectable images in both datasets.

4 Providing labeled data for training and testing

For training and evaluating the networks, a sufficient number of annotated reference samples were required. Therefore, a simple, mouse-based computer method programmed in C# was used (Microsoft, 2024). Image sections of 128×128 pixels each were selected from various original images, such as

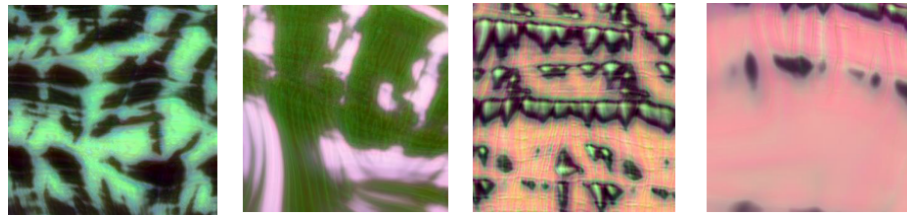


Figure 7. Results of dataset 1 for values of $\text{Kimg} \geq 1040$.

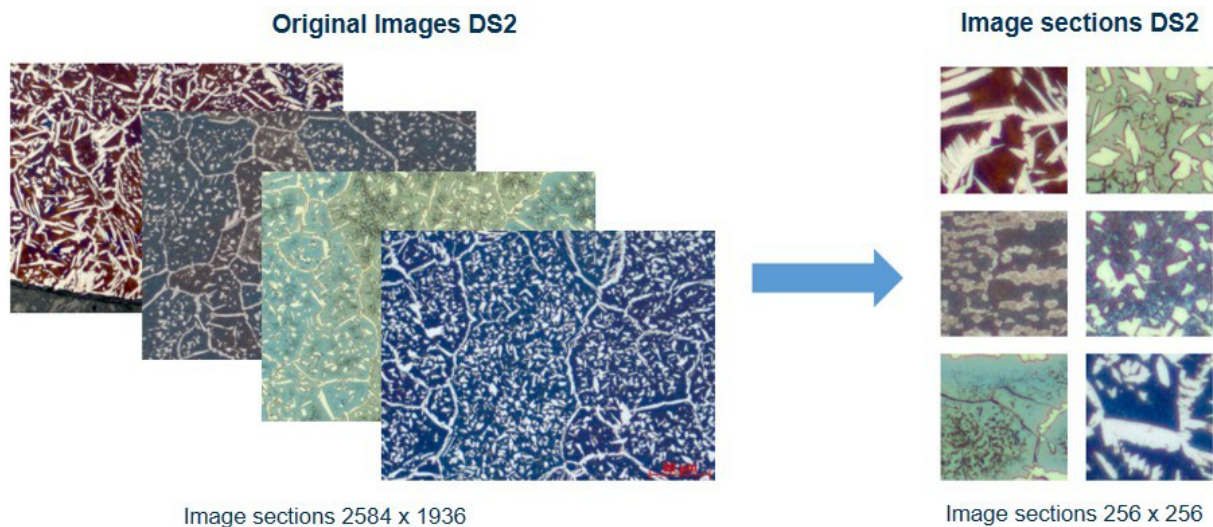


Figure 8. Images and image sections of dataset 2.

those in Fig. 3, or from augmented images. The image size of 128×128 pixels was chosen because it represents an optimal compromise between sufficient recognizability of the corresponding phase structures and practical manual annotation in the labeling process while also being efficient for training a convolutional neural network (CNN) in terms of memory requirements and computational effort. Austenite regions were identified manually/visually by moving the mouse pointer (see Fig. 10). All pixels within these areas were assigned the status “austenite”, and all pixels outside these areas were assigned the status “ferrite”. This approach is similar to the well-known but non-computerized point-counting method ASTM E562-19 (ASTM International, 2019). However, it offers the advantage of classifying a larger number of pixels in a single pass.

5 Use of neural networks

To solve the problem, the use of various neural networks was investigated (Basler, 2021). Two perceptrons, as well as a convolutional neural network (CNN, see Fig. 11), and a modification of it (Khan et al., 2018) were used. Networks built with the Tensorflow package (Abadi et al., 2015) under the Anaconda Platform (Anaconda Software Distribution, 2024)

were used for each of these. Programming, including the network configuration and access to Tensorflow, was done in the Python programming language Python 3.10.15 (Anaconda Software Distribution, 2024; Python Software Foundation, 2024).

The perceptrons were networks with 25×25 inputs for a correspondingly large image section and only two outputs for the classes “Fe” and “Au”. During a network pass, the class membership is calculated for the middle pixel of the input. The $25 \times 25 - 1$ pixels of the input are included in the calculation. A network with one intermediate layer and a network with two intermediate layers were used. The layers are fully connected. A different approach was pursued with CNNs. These are deep networks with convolutional, pooling, and upscaling layers. They have 128×128 inputs for a correspondingly large image section and 128×128 outputs for the class assignments of all inputs. Thus, 128×128 class assignments are calculated in one network run. For both types of networks, the same hardware platform with the following performance parameters was used to execute the program:

- AMD Ryzen 5900HX, clock speed 3.3 GHz, 16 threads;
- NVIDIA GeForce RTX 3060 graphics processor, 6 GB graphics memory.

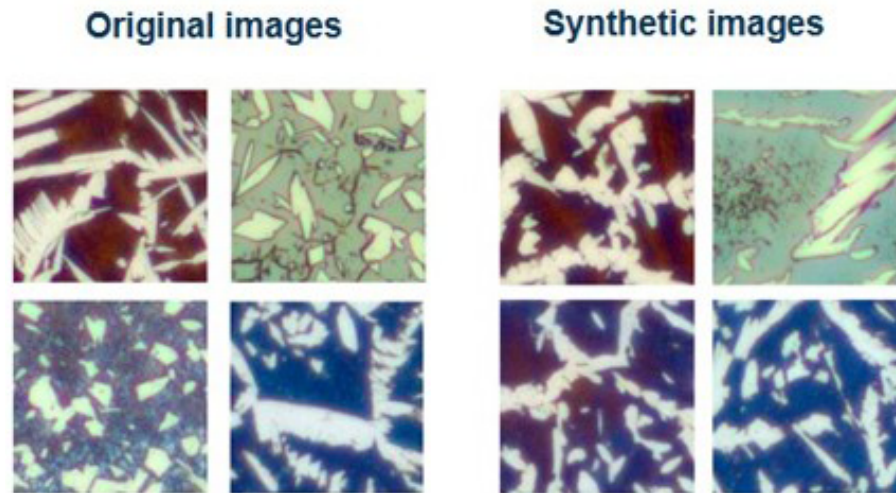


Figure 9. Results of image augmentation, dataset 2. GAN used: StyleGAN3 (Karras et al., 2021). Image size: 256×256 . Calculation on a computing cluster at TU Ilmenau, official PyTorch implementation pre-trained model LHQ-256 (Pinkney, 2019).

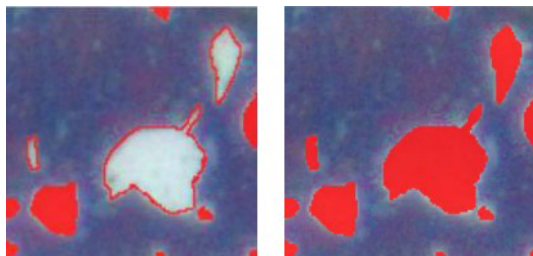


Figure 10. Providing labeled data for training and validation. Left panel: hovering over the austenite regions with the mouse on the PC. Right panel: pixels of the encircled regions are set to “Au” (austenite), and all others are set to “Fe” (ferrite).

6 Training

For training, a number of labeled patterns were selected, applied to the networks one after the other over several epochs, and the networks were trained by specifying the labels. The learning progress was monitored after each epoch by determining the parameter accuracy (Acc) and intersection over union (IoU) (Huynh, 2023) and recorded in learning curves. Various network variants, as well as real data, augmented data, and combinations of both data types, were used. For the perceptrons, feature vectors of 25×25 pixels were applied to the network, and, for the convolutional networks, 128×128 pixels were applied.

7 Tests and results

The next step was evaluation. The neural networks were applied to labeled data that had not been used for training. The Acc and IoU parameters were again used as criteria. This

was performed for various network variants and with variants trained with different datasets. The results of selected variants are shown in Table 1. As can be seen, the accuracy results are all between 97 % and 98 %, and the IoU results are between 90 % and 93 %, which is in line with expectations.

The best results were achieved with the modified convolutional neural network and a training dataset with blue and green samples. While CNN1 showed higher accuracy and IoU on the test data than on the training data, suggesting a possible overfitting to the training data, this behavior was reduced in the improved model (CNN2), indicating a more stable generalization. However, with the addition of augmented data and, finally, an additional color variant, the results improved. The incremental increases appear to be marginal. Even small improvements in the IoU in the high percentage range represent a significant performance increase for the model as they reflect more precise segmentation and more robust generalization. It is also likely that further optimization could lead to an increase of over 99 %. While the results regarding achievable accuracy are roughly similar for all network variants, there are significant differences in terms of execution speed. This is approximately 4 times faster for CNNs compared to perceptrons, which is attributable to a more efficient network structure.

8 Conclusion

The presented work builds the foundation for the automatic determination of phase fractions in duplex stainless steel. The application of neural networks to determine the phase fraction has been functionally demonstrated. Accuracy values of 97 % and 98 % (Acc) and between 90 % and 93 % (IoU) were achieved, which is in line with expecta-

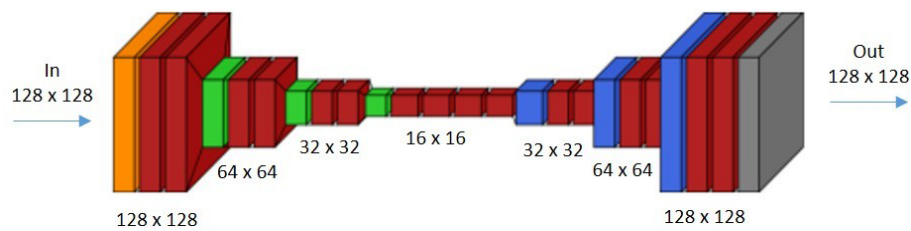


Figure 11. Convolutional neural network used in the project. Orange: input layer; red: convolutional layers; green: pooling layers; blue: upscaling layers; gray: output layer.

Table 1. Results of training and testing for different network variants and datasets.

Network	Dataset	Number of training data				2560 × 1280 px		Training		Test	
		Total	Original	Augmented		Sample test image		Acc	IoU	Acc	IoU
				Non-destr.	GAN	Ground truth austenite % = 26.504					
						Network runtime/s	Predicted austenite %				
Per1	blue	625	625	0	0	72	26.895	0.968	0.898	0.971	0.899
Per2	blue	625	625	0	0	102	27.562	0.972	0.894	0.971	0.898
CNN1	blue	280	280	0	0	26.3	26.891	0.976	0.917	0.970	0.896
CNN2	blue	280	280	0	0	29.3	27.036	0.971	0.903	0.975	0.911
CNN2	blue	4480	560	1960	1960	29.3	26.506	0.974	0.912	0.977	0.918
CNN2	blue + green	6496	812	3724	1960	29.3	26.505	0.9813	0.9285	0.9847	0.933

Per1/2: perceptron with one and two intermediate layers. Dataset refers to 25 × 25 pixels (perceptron) or 128 × 128 pixels (CNN). CNN1: as shown in Fig. 10; CNN2: modified version of CNN1. Blue/green: sections of images as shown in Fig. 3, second row, first and second images from the left.

tions. Future work will explore additional application areas in metallography and ceramography, further increasing accuracy and expanding the range of materials to be examined. In particular, foldable neural networks offer great potential in this area. Other possible future applications include detection of grain and phase boundaries; the detection of alloy components; and the detection of cavities, including in X-ray, CT, or ultrasound images.

Code availability. The code developed for this study is not publicly available at this time, but can be obtained from the corresponding author upon reasonable request.

Data availability. The datasets generated and analyzed during this study are not publicly available at this time but are available from the corresponding author upon reasonable request.

Author contributions. LK, together with TS, designed the concept for the AI. LK also programmed the AI. JW produced the metallography samples and described the production process. GP conducted and described the investigations into the application of data augmentation to the metallography samples. TS edited the complete paper and, together with LK, designed the concept for the AI.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements. The investigations into the metallography of duplex steels were carried out by the Günter-Köhler-Institut für Fügetechnik und Werkstoffprüfung GmbH. They were funded by the Ministry of Economic Affairs and Climate Protection (funding code 49MF210091). The investigations into data augmentation using Generative Adversarial Networks (GANs) were carried out within the framework of the ThurAI and ProKI projects at the Department of Quality Assurance and Industrial Image Processing at Technische Universität Ilmenau. The ThurAI project was funded by the Free State of Thuringia and co-financed by the European Union within the framework of the European Regional Development Fund (ERDF). The ProKI project was funded by the Federal Ministry of Education and Research (BMBF) within the program “Future of Value Creation – Research on Production, Services

and Work” (funding code 02P22A040) and was supervised by the Project Management Agency Karlsruhe (PTKA).

The authors gratefully thank the funding authorities for their financial support.

Financial support. This research has been supported by the Federal Ministry for Research and Technology (grant no. 49MF210091).

Review statement. This paper was edited by Sebastian Wood and reviewed by two anonymous referees.

References

- Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattemberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (Version 2.10.0), TensorFlow, <https://www.tensorflow.org/> (last access: 18 September 2024), 2015.
- Anaconda Software Distribution: Anaconda Documentation (Version 2.5.2), Anaconda Inc., <https://docs.anaconda.com/> (last access: 18 September 2024), 2024.
- ASTM International: Standard Test Method for Determining Volume Fraction by Systematic ManuPoint Count (ASTM E562-19), ASTM International, <https://doi.org/10.1520/E0562-19>, 2019.
- Basler, D.: Neuronale Netze mit C# programmieren – Mit praktischen Beispielen für Machine Learning im Unternehmenseinsatz, Carl Hanser Verlag GmbH Co., KG, ISBN 978-3-446-46426-1, 2021.
- Beraha, E.: Farbätzung für die Erkennung von Phosphiden, Karbiden und Nitriden in Eisen, Stahl, Werkzeugstahl, rostfreiem Stahl und hitzebeständigen Legierungen, in: Prakt. Metallogr., vol. 5, Walter de Gruyter, Berlin, Boston, 501–508, <https://doi.org/10.1515/pm-1968-050904>, 1968.
- Cogswell, M., Ahmed, F., Batra, D., Grishick, R. B., and Zitnick, C. L.: Reducing overfitting, in deep networks by decorrelation representations, in: Int. Conf. on Learning Representation, 19 November 2015, arXiv [preprint], <https://doi.org/10.48550/arXiv.1511.06068>, 2015.
- Folkhard, E.: Metallurgie der Schweißung nichtrostender Stähle, Springer Verlag, Vienna, New York, ISBN 978-3211818039, 1983.
- Goodfellow, I. J., Bengio, Y., Courville, A., Mirza, M., Ozair, S., Pouget-Abadie, J., Warde-Farley, D., and Xu, B.: Generative adversarial nets, in: Proc. of the 27th Int. Conf. on Neural Information Processing Systems – vol. 2, MIT Press, Montreal, Canada, 2672–2680, <https://doi.org/10.48550/arXiv.1406.2661>, 2014.
- Huynh, N.: Understanding Evaluation Metrics in Medical Image Segmentation, Version: 202, <https://medium.com/mastering-data-science/understanding-evaluation-metrics-in-medical-image>, (last access: 18 September 2024), 2023.
- Karras, T., Aila, T., and Laine, S.: Supplemental material: A style-based Generator architecture for generative adversarial networks, arXiv [preprint], <https://doi.org/10.48550/arXiv.1812.04948>, 2019.
- Karras, T., Aila, T., Aittala, M., Laine, S., Härkönen, E., Hellsten, J., and Lehtinen, J.: Alias-Free Generative Adversarial Networks, arXiv [preprint], <https://doi.org/10.48550/arXiv.2106.12423>, 2021.
- Khan, S., Bennamoun, M., Rahmani, H., and Shah, S. A. A.: A Guide to Convolutional Neural Networks for Computer Vision, Morgan Claypool Publishers, London, ISBN 978-1-681-73022-6, <https://doi.org/10.1109/LGRS.2018.2802944>, 2018.
- Microsoft: C# (Version 12), Microsoft, <https://learn.microsoft.com/en-us/> (last access: 18 September 2024), 2024.
- Otsu, N.: A Threshold Selection Method from Gray-Level Histograms, in: IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66, <https://doi.org/10.1109/TSMC.1979.4310076>, 1979.
- Pinkney, J.: Modell LHQ-256, <https://www.justinpinkney.com/> (last access: 18 September 2024), 2019.
- Python Software Foundation: Python (Version 3.10.15), <https://www.python.org/> (last access: 18 September 2024), 2024.
- Ulrich, S., Jahn, S., and Schaaf, P.: Influence of the process parameters on the corrosion behavior of quality 2205 duplex stainless steel in laser beam welding with highest focusability, in: 72nd IIW Annual Assembly and International Conference 2019, 7–12 July 2019, Bratislava, https://doi.org/10.7449/2019/MST_2019_1018_1026, 2019.
- Wunsch, L., Anding, K., Liu, K., Notni, G., and Polte, G.: Data augmentation for solving industrial recognition tasks with underrepresented defect classes, Acta IMEKO, 12, 41, <https://doi.org/10.21014/actaimeko.v12i4.1320>, 2023.