

Article

Training Tricks for Steel Microstructure Segmentation with Deep Learning

Xudong Ma¹ and Yunhe Yu^{2,*}

¹ State Key Laboratory of Rolling and Automation, Northeastern University, Shenyang 110819, China; 2310186@stu.neu.edu.cn

² Shagang School of Iron and Steel, Soochow University, Suzhou 215137, China

* Correspondence: yhyu@suda.edu.cn; Tel.: +86-136-1404-2499

Abstract: Data augmentation and other training techniques have improved the performance of deep learning segmentation methods for steel materials. However, these methods often depend on the dataset and do not provide general principles for segmenting different microstructural morphologies. In this work, we collected 64 granular carbide images (2048×1536 pixels) and 26 blocky ferrite images (2560×1756 pixels). We used five carbide images and two ferrite images and derived from them the test set to investigate the influence of frequently used training techniques on model segmentation accuracy. We propose a novel method for quickly building models that achieve the highest segmentation accuracy for a given dataset through combining multiple training techniques that enhance the segmentation quality. This method leads to a 1–2.5% increase in mIoU values. We applied the optimal models to the quantization of carbides. The results show that the optimal models achieve the smallest errors of 5.39 nm for the mean radius and 29 for the total number of carbides on the test set. The segmentation results are also more reasonable than those of traditional segmentation methods.

Keywords: steel microstructure; segmentation; transfer learning; data augmentation; receptive field



Citation: Ma, X.; Yu, Y. Training Tricks for Steel Microstructure Segmentation with Deep Learning. *Processes* **2023**, *11*, 3298. <https://doi.org/10.3390/pr11123298>

Academic Editors: Weisen Zheng, Wei Lv, Wanqiang Liu, Li Yang and Feng Wang

Received: 23 October 2023
Revised: 15 November 2023
Accepted: 17 November 2023
Published: 26 November 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Quantitative microstructure analysis is at the heart of materials engineering and design [1]. Microscopy image segmentation is usually the first and most difficult step in quantifying material structure. Traditionally, it is often processed manually or using expensive equipment. With the development of practical stereology [2] and digital techniques, theoretical studies performed with simple image processing software, such as image thresholding, have attracted much attention. Martyushev et al. [3,4] present the software that can automate the analysis and furthermore determine upper and lower boundaries for the intensity levels of the corresponding phases. Stuckner et al. [5] present a Python package for complex multiphase material analysis. These methods are accurate and reproducible but are not stable to small changes in imaging or sample conditions, and the complex theoretical foundations hamper their further extension. Therefore, the development of a simple and convenient segmentation method is a pressing problem for us.

Deep learning [6] has achieved impressive performance in visual domains such as autonomous driving [7,8] and healthcare [9,10], creating new opportunities for experts to utilize images directly in their AI applications. Unlike traditional complex theoretical methods, deep learning fits a large number of parameters through mimicking the human learning process. It has high robustness and is easy to apply to material classification and quantization processes, which has garnered significant attention. For example, Cui et al. [11] proposed a deep learning method for additive manufacturing quality inspection with a high practical value. Additionally, the proposal of fully convolutional neural networks (FCNs) has opened up new avenues for pixel-level segmentation [12–14].

Ma et al. proposed a symmetric 3D information segmentation method based on Al–La alloy micrographs that enables segmentation of high-resolution images. However, the study used artificially (subjectively) generated labeled data to train the network, which led to some errors. To address this challenge and achieve more accurate segmentation for complex organizations, Shen et al. [13] proposed a deep learning method based on EBSD data to classify and quantify stainless steels and QP steels using only conventional SEM images, while achieving robustness of segmentation effects for different forming qualities and magnifications. S. Breumier et al. [14] developed a U-Net model for the segmentation of ferrite, bainite, and martensite using EBSD data. The overall accuracy of the model was 92%, and the accuracy of each sample varied between 86.6% and 95% based on its microstructure complexity.

While model accuracy has steadily improved, most semantic segmentation algorithms still lack a systematic theoretical study due to the difficulty of collecting steel material image datasets and the lack of transparency in convolutional neural network (CNN) decision making [15]. Additionally, training strategies, such as transfer learning, data preprocessing, and loss function optimization, play an important role. In recent years, numerous improvements have been proposed, but they are often limited to a specific method for a specific dataset and lack a generalizable model guide. Therefore, establishing a comprehensive model training system to fully exploit the potential of existing deep learning segmentation models is an urgent need.

In this paper, we investigate methodological techniques to improve the segmentation performance of deep learning models for steel material image datasets without increasing model size or computational complexity. We consider common training techniques such as transfer learning and data augmentation. Transfer learning improves a target learner's performance in a current domain through transferring knowledge from different but related domains [16]. It utilizes previously collected data to its full potential. Data augmentation prevents model overfitting through simulating possible changes in real data, extracting more generalized information and features from small datasets [17]. We also investigate reducing the input image size to improve the model's receptive field, given the importance of microstructures' multi-scale nature for model training. We evaluate our techniques on multiple network architectures and datasets to provide effective training guidelines for building models with optimal segmentation accuracy.

The remainder of this paper is organized as follows: Section 2 presents the datasets and methods used in the study. Section 3 presents the segmentation results of the baseline training model with and without training techniques. Section 4 discusses the process of building an optimal model for a given dataset, and compares the quantization results of the optimal models with traditional methods to demonstrate the superiority of deep learning.

2. Materials and Methods

The overall framework is shown in Figure 1. First, we will establish a baseline model training procedure. Then, we will investigate the impact of common training techniques on segmentation accuracy to prepare for constructing the most accurate model and microstructure quantization. This section describes the dataset construction process and the specific implementation of the baseline model and training techniques.

2.1. Dataset Description

To develop training guidelines for the segmentation of different material organizational forms in SEM images of steel, we collected 90 images of Fe-0.2C-1.35Mn-2.5Cr-1.5Si alloy subjected to different heat treatments. The data were divided into two parts: 64 high-magnification (20,000×) SEM images of carbide precipitation and 26 ordinary (5000×) SEM images of ferrite, martensite, and residual carbides. Both parts underwent austenitizing, tempering, and annealing, but the former had a shorter tempering time. To account for the differences in quantity and texture between the two parts, we separated them into two datasets: the carbide dataset with categories of carbide and matrix, and the ferrite

dataset with categories of ferrite and matrix (martensite and residual carbide). A materials science expert labeled the two datasets, as shown in Figure 2.

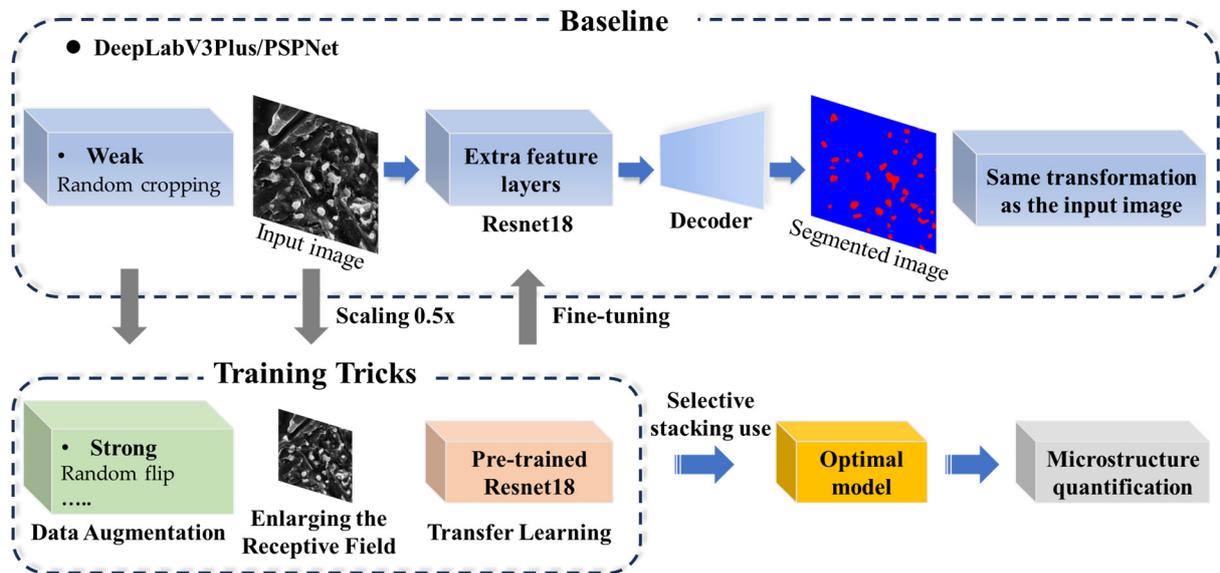


Figure 1. Experimental procedure of this study.

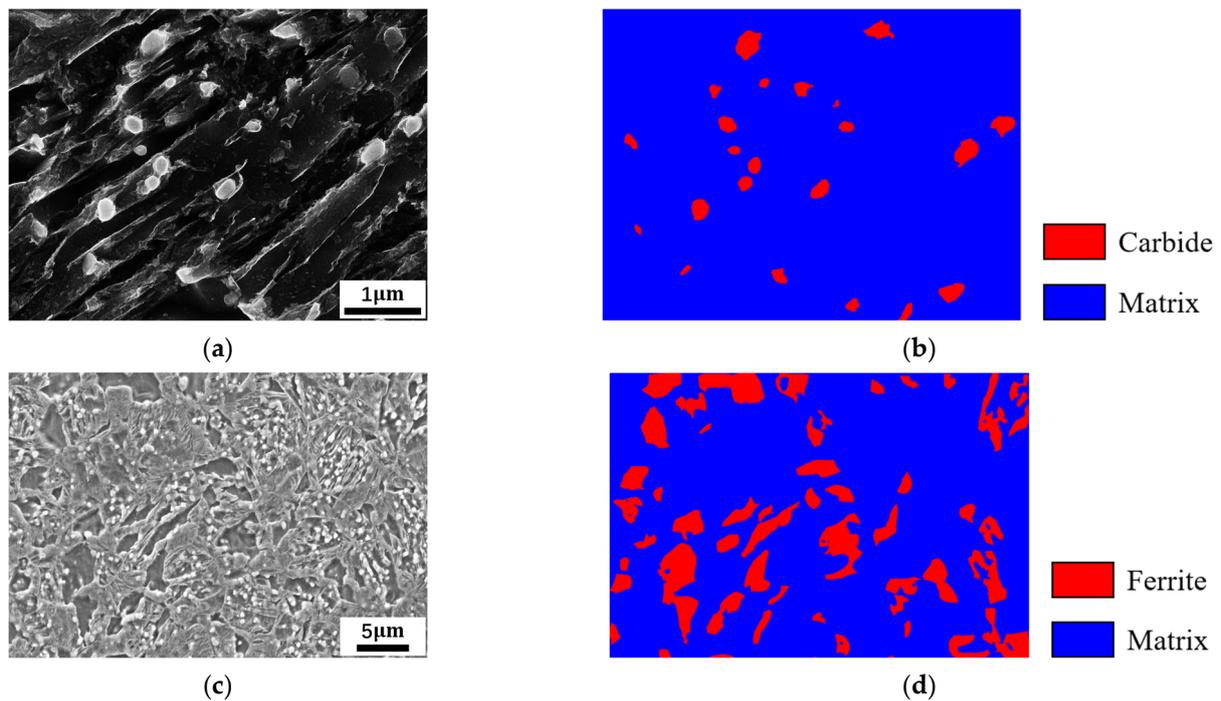


Figure 2. Typical images: (a) the carbide dataset image; (c) the ferrite dataset image; (b) labeled image from (a) input image; (d) labeled image from (c) input image.

For the carbide dataset, the training set consists of 59 images of 2048×1536 pixels as input and 59 labeled images as output. The test set consists of the remaining five images. For the ferrite dataset, the training set consists of 24 images of 2560×1756 pixels as input and 24 labeled images as output. The test set consists of one image from the same process as the training set and one image from a different process.

We did not augment the datasets offline but used online augmentation. See Section 2.2 for more details.

2.2. Baseline Training Procedure

In this study, we trained two widely used semantic segmentation architectures, PSP-Net [18] and DeepLabV3Plus [19], with ResNet18 as the encoder using PyTorch [20] and the Segmentation Model Library (MMSegmentation) [21]. We used Focal Loss [22] with alpha and gamma of 0.25 and 2, respectively, as the loss function because it has a stable gradient, is more robust for unbalanced categories, and resembles the true goal of maximizing intersection over union (IoU) [23]. We used the standard SGD optimizer with an initial learning rate of 1×10^{-2} and a batch size of eight. To accelerate convergence and avoid oscillating around the optimal point late in training, we used a polynomial decay strategy for the learning rate. After 8000 training generations, the learning rate decayed to 0, and the model stopped training.

In the data loading stage, we first resized the input images to a specified size and decoded them into 32-bit floating-point raw pixel values in the range [0, 255]. Then, we randomly cropped each image into a square region with a width and height of 256 pixels, ensuring that no single category occupied more than 75% of the cropped region. Finally, we normalized all regions through subtracting 123.675, 116.28, and 103.53, respectively, and dividing by 58.395, 57.12, and 57.375, respectively.

In the validation stage, we used the sliding window model to perform inference. Specifically, we slid a fixed window from left to right and from top to bottom over the test image. At each slide, we normalized the current window in the same way as during training, then used the segmentation model to perform inference. We repeated this process until we had covered the entire image. Finally, we stitched the inference results together. To mitigate edge cracking, we set the sliding window size to 256×256 and the moving step size to 128×128 . To further improve segmentation accuracy, we adopted the “test set data augmentation” method, in which we scaled the test image to different multiples and performed inference on each scaled image. We then averaged the output results to obtain the final prediction. We trained and validated the model on a system equipped with an Intel(R) Xeon(R) Gold 6271 CPU (16 cores) and an NVIDIA Tesla P100 (16 GB video memory) graphics card.

2.3. Training Tricks

2.3.1. Transfer Learning

Larger datasets generally produce better deep learning models [24,25]. However, materials scientists face the challenge of small sample problems, as collecting and labeling data typically requires manual effort, and factories and laboratories only have specific steel grades.

Transfer learning is a powerful machine learning technique that enables knowledge sharing between related domains [26–28]. However, for the materials domain, there is still no clear standard for selecting and transferring which dataset due to the lack of large, high-quality labeled datasets. Typically, we pre-train on the ImageNet dataset [29]. Given the texture difference between natural and microscopic image data, we only transfer the encoder of the segmentation model, i.e., for the baseline training model, we load the model parameters provided by the model library [21] for the encoder part (ResNet18) and do not make any changes to other parts.

2.3.2. Strong Data Augmentation

Data augmentation, a popular approach for solving data scarcity, involves expanding the size of a dataset using heuristics or synthetic samples [30]. It can be divided into two main categories: transformations, which involve applying multiple operations to existing data, such as cutout [31] and random erasing [32], and generative models, which involve using a generated model to generate new data, such as variational autoencoders (VAEs) [33].

The original data augmentation method only includes random cropping. We introduce several additional data augmentation methods based on this, such as random flip with

0.5 probability; 90 degrees clockwise rotation with 0.5 probability; random cutout with 0.5 probability, which masks five areas of 5×5 pixels, with masked areas set to zero; and finally, applying photometric distortion (random brightness and random contrast). We refer to the original augmentation method as weak data augmentation. We refer to this new augmentation as strong data augmentation.

2.3.3. Enlarging the Receptive Field

The receptive field of a CNN model is the size of the region in the input image that is mapped to a pixel in the feature map output by a given layer. The larger the receptive field, the greater the model's scope of perception of the image's context, which can lead to a better understanding of the semantic associations between pixels. For example, in the semantic segmentation of images containing multiple objects of different categories, a larger receptive field can provide the model with a wider range of contextual information, helping it to make more accurate classification decisions. This is especially important in the materials domain, where observations at different scales have different physical meanings, such as phases and grain boundaries. A too-small receptive field can split the original features into small fragments and lose the original physical meaning, which can affect the segmentation accuracy. Therefore, increasing the receptive field is a necessary operation for semantic segmentation in the materials domain. However, common operations to increase the receptive field, such as pooling and increasing the number of layers, can increase the computational complexity of the model and lose a lot of favorable information.

To increase the receptive field of a CNN model without modifying its architecture or losing favorable information, we propose a simple and effective approach: scaling the original image size to 0.5 times during image loading.

2.4. Evaluation Indicators and Quantitative Analysis Process

To evaluate the performance of our semantic segmentation models, we will use two standard metrics: the intersection over union (IoU) for a specific microstructure type (carbide or ferrite) and the mean intersection over union (mIoU) for all microstructure types. These metrics are defined in Equations (1) and (2).

$$IoU = \frac{TP}{TP + FP + FN} \quad (1)$$

$$mIoU = \frac{1}{k} \sum_{i=0}^{k-1} IoU_i \quad (2)$$

Here, k represents the number of categories (which is two in this study). FP denotes false positives (samples predicted to be positive but are actually negative), FN denotes false negatives (samples predicted to be negative but are actually positive), TP denotes true positives (correctly categorized positive samples), and TN denotes true negatives (correctly categorized negative samples).

An automatic quantitative analysis of microstructure is established using the OpenCV software package (version 4.6.0.) based on the existing trained segmentation model. The quantitative process is as follows:

The target SEM image is directly fed into the trained segmentation model for pixel-level classification of organizations using the window sliding mode, eliminating the need to cut sub-images beforehand. The morphological information of the organizations, such as average carbide radius and number of carbides, can be quickly calculated using the OpenCV method based on the obtained output image.

3. Results

3.1. Baseline Training

We evaluated the validation results of the aforementioned baseline models on both datasets. Table 1 shows that DeepLabV3Plus and PSPNet achieved mIoU values of 80.38%

and 81.14% on the carbide dataset, respectively, and 80.35% and 80.32% on the ferrite dataset, respectively. These mIoU values are higher than 80%, suggesting that both models have high segmentation accuracy. It should be noted here that unlike Shen et al. [13], our evaluations are performed on the original large-size images, rather than the small images after slicing.

Table 1. Validation accuracies of different models trained with our “tricks”. The number in the table represents the size of the mIoU value (%).

Dataset	Method	Baseline	Transfer Learning Gain	Strong Data Augmentation Gain	Enlarging the Receptive Field Gain
Carbide	DeepLabV3Plus	80.38	+0.64	+1.35	−0.37
	PSPNet	81.14	−1.06	+1.19	+0.15
Ferrite	DeepLabV3Plus	80.35	+0.75	−0.79	+0.87
	PSPNet	80.32	+0.68	+0.25	+0.24
	Average gain		+0.25	+0.5	+0.22

Figures 3b, 4b, 5b and 6b show the validation results of the baseline training models. White pixels indicate true positive predictions, black pixels indicate true negative predictions, green pixels indicate false positive predictions, and pink pixels indicate false negative predictions. As shown in Figures 3a, 4a, 5a and 6a, these raw input images contain ambiguous boundaries and numerous interfering regions, such as bright boundaries in the matrix of the carbide dataset. Therefore, it is difficult to perform fast and accurate segmentation. However, the semantic segmentation models solve this problem, and most of the carbides and ferrites are recognized. A small number of matrix regions are misidentified as positive samples, such as the portion in the orange box. This is understandable, given the high similarity in texture and brightness between it and normal carbides, and the fact that the baseline training models lack a priori knowledge of the materials science.

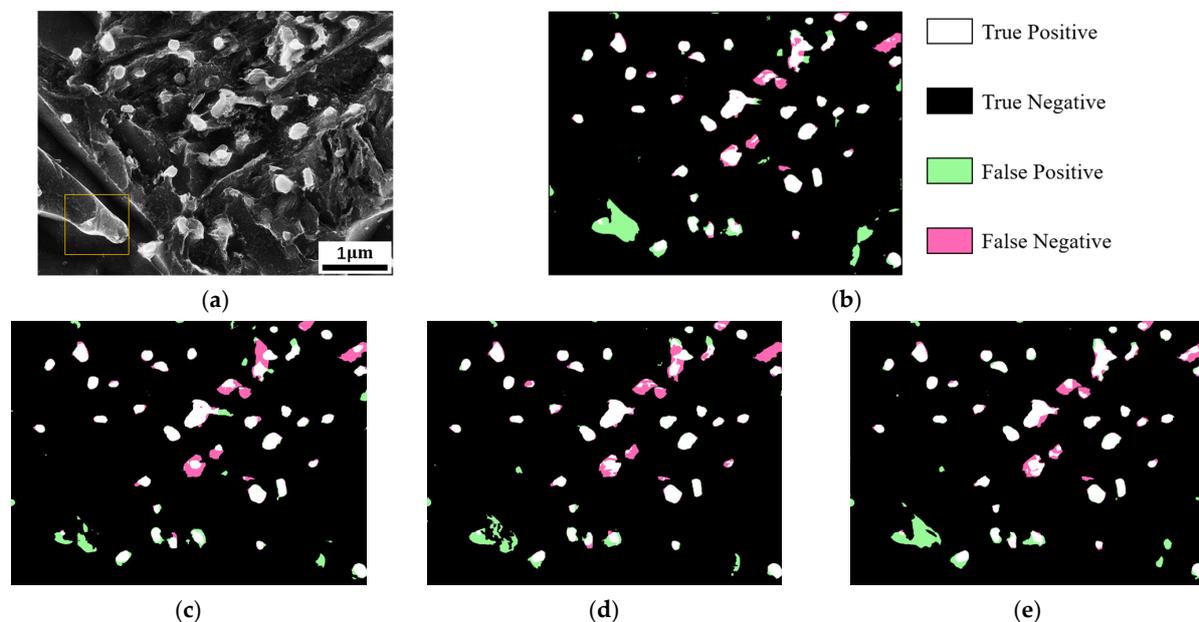


Figure 3. Segmentation results of DeepLabV3Plus for the carbide dataset. (a) The original input image. (b) The segmentation image of the baseline training program. (c–e) The segmentation images after applying the techniques of transfer learning, strong data augmentation, and receptive field enlargement, respectively. White pixels indicate true positive predictions, black pixels indicate true negative predictions, green pixels indicate false positive predictions, and pink pixels indicate false negative predictions.

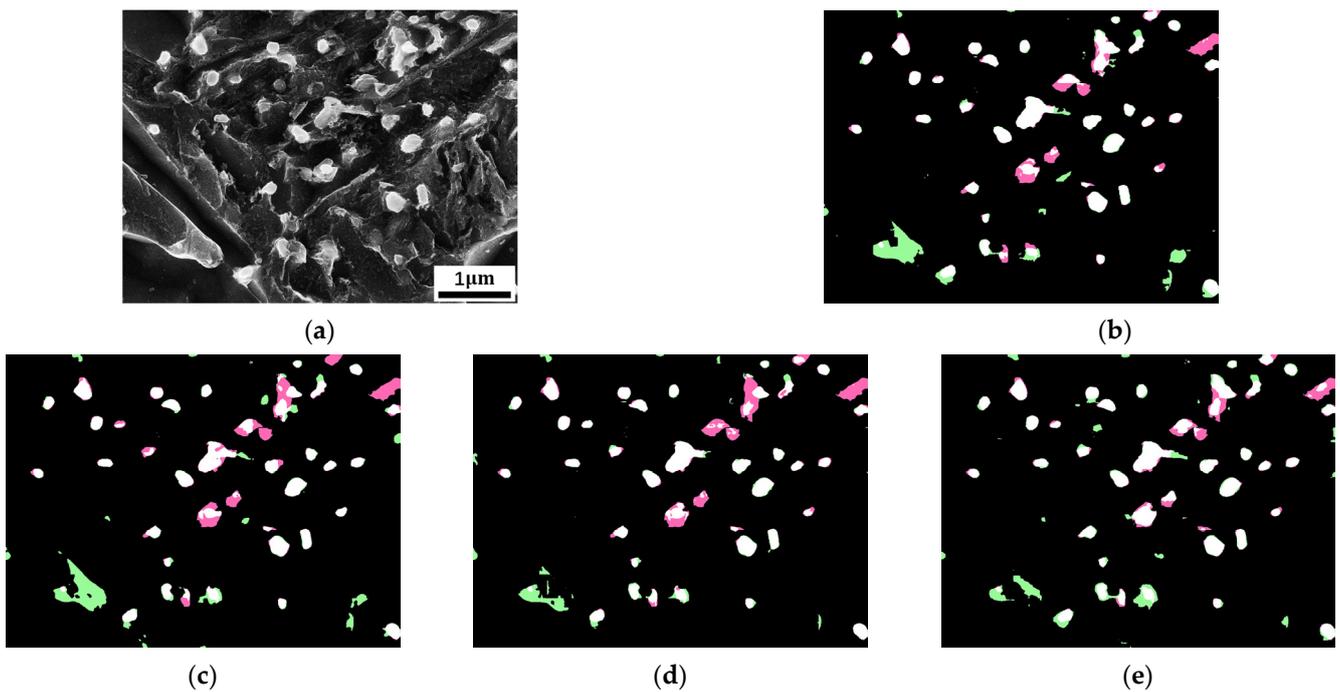


Figure 4. Segmentation results of PSPNet for the carbide dataset. (a) The original input image. (b) The segmentation image of the baseline training program. (c–e) The segmentation images after applying the techniques of transfer learning, strong data augmentation, and receptive field enlargement, respectively. The color scheme is consistent with Figure 3.

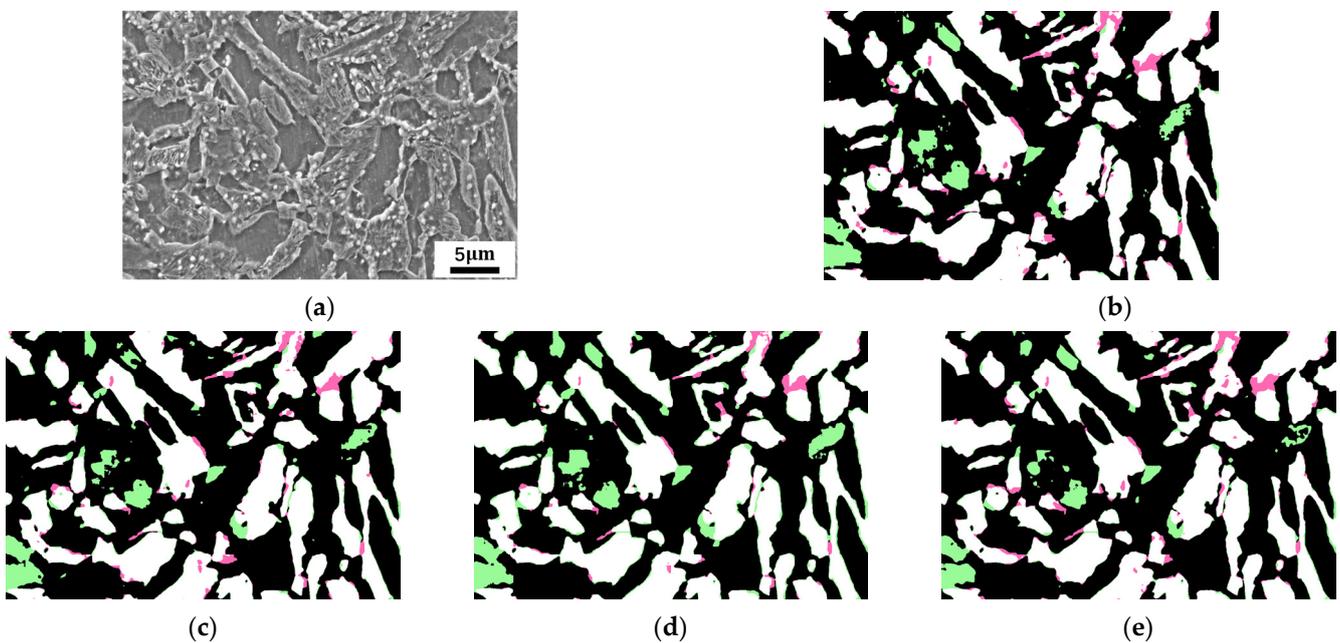


Figure 5. Segmentation results of DeepLabV3Plus for the ferrite dataset. (a) The original input image. (b) The segmentation image of the baseline training program. (c–e) The segmentation images after applying the techniques of transfer learning, strong data augmentation, and receptive field enlargement, respectively. The color scheme is consistent with Figure 3.

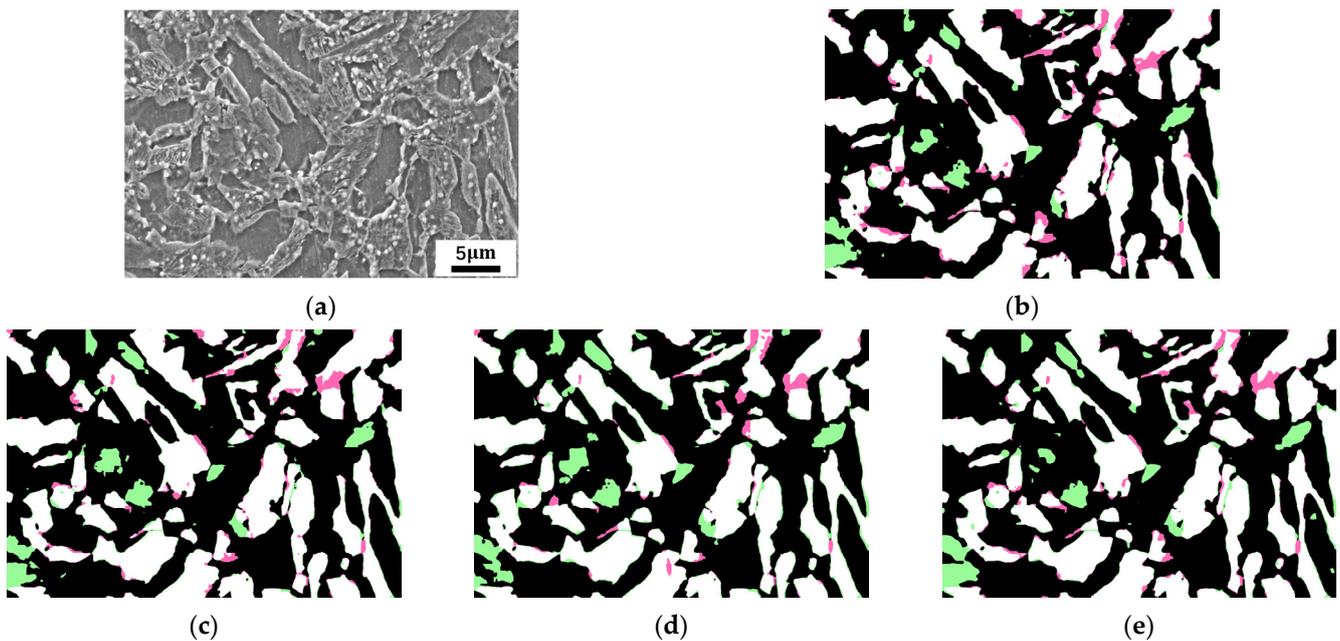


Figure 6. Segmentation results of PSPNet for the ferrite dataset. (a) The original input image. (b) The segmentation image of the baseline training program. (c–e) The segmentation images after applying the techniques of transfer learning, strong data augmentation, and receptive field enlargement, respectively. The color scheme is consistent with Figure 3.

3.2. Baseline Training Procedure with Training Trick

Table 1 shows the evaluation results of the baseline training models with training tricks for the two datasets. For the carbide dataset, the DeepLabV3Plus model achieves a higher mIoU with transfer learning and strong data augmentation than the baseline model, but a lower mIoU with smaller input images. These results suggest that transfer learning and strong data augmentation are beneficial training techniques for the DeepLabV3Plus model on this dataset. Similarly, the PSPNet model benefits from enlarging the receptive field and strong data augmentation. For the ferrite dataset, the DeepLabV3Plus model benefits from transfer learning and enlarging the receptive field, while the PSPNet model benefits from transfer learning, strong data augmentation, and enlarging the receptive field.

Figures 3c–e and 4c–e show the segmentation results of the DeepLabV3Plus and PSPNet models for the carbide dataset with the addition of a training technique. These figures reveal that the misclassified regions, especially the false positive regions, become smaller with the addition of a beneficial technique, demonstrating the improved ability of the baseline models to recognize other bright interfaces in the matrix. In contrast, after adding an unfavorable technique, the misclassified regions become larger, which is consistent with the trend of mIoU. Figures 5c–e and 6c–e show the segmentation results of the models on the ferrite dataset, and the transformation is not obvious after adding training techniques.

Based on the above results, transfer learning, strong data augmentation, and enlarging the receptive field have improved model prediction accuracy in most cases, indicating that these techniques can play a positive role without significantly increasing computational complexity.

4. Discussion

4.1. Building the Optimal Segmentation Model

Constructing a highly accurate model for subsequent quantization is a challenge we must address. Based on the findings in Section 4.2, we will gradually stack beneficial training techniques onto the baseline model to explore the possibility of constructing a model with the highest mIoU value.

Tables 2–5 show that gradually stacking beneficial training techniques onto the baseline training models gradually increases the IoU and mIoU values. Specifically, the mIoU values of DeepLabV3Plus and PSPNet increased by 2.43% and 1.36%, respectively, on the carbide dataset, and by 1.91% and 1.45%, respectively, on the ferrite dataset. The highest model accuracy was achieved when all beneficial training techniques were superimposed. This suggests that to train an optimal model, we can simply explore the effect of each training technique on the baseline model and superimpose the beneficial training techniques. Therefore, if we set the number of training techniques to n , we only need to train $n + 2$ models.

Table 2. Validation accuracy of DeepLabV3Plus on the carbide dataset through stacking beneficial training techniques. The IoU values of carbide and ferrite are expressed as $\text{IoU}_{\text{carbide}}$ and $\text{IoU}_{\text{ferrite}}$, respectively.

Method	Transfer Learning	Strong Data Augmentation	Enlarging the Receptive Field	$\text{IoU}_{\text{carbide}}\%$	mIoU%
DeepLabV3Plus				63.65	80.38
DeepLabV3Plus	✓			64.7	81.02
DeepLabV3Plus	✓	✓		67.88	82.81

A check mark indicates that the corresponding training technique has been applied.

Table 3. Validation accuracy of PSPNet on the carbide dataset through stacking beneficial training techniques.

Method	Transfer Learning	Strong Data Augmentation	Enlarging the Receptive Field	$\text{IoU}_{\text{carbide}}\%$	mIoU%
PSPNet				64.95	81.14
PSPNet		✓		67.11	82.33
PSPNet		✓	✓	67.52	82.5

A check mark indicates that the corresponding training technique has been applied.

Table 4. Validation accuracy of DeepLabV3Plus on the ferrite dataset through stacking beneficial training techniques.

Method	Transfer Learning	Strong Data Augmentation	Enlarging the Receptive Field	$\text{IoU}_{\text{ferrite}}\%$	mIoU%
DeepLabV3Plus				71.19	80.35
DeepLabV3Plus	✓			72.32	81.1
DeepLabV3Plus	✓		✓	73.75	82.26

A check mark indicates that the corresponding training technique has been applied.

Table 5. Validation accuracy of PSPNet on the ferrite dataset through stacking beneficial training techniques.

Method	Transfer Learning	Strong Data Augmentation	Enlarging the Receptive Field	$\text{IoU}_{\text{ferrite}}\%$	mIoU%
PSPNet				70.83	80.32
PSPNet	✓			72.06	81
PSPNet	✓	✓		72.28	81.09
PSPNet	✓	✓	✓	73.1	81.77

A check mark indicates that the corresponding training technique has been applied.

4.2. Comparison with the Traditional Binary Segmentation Method

We used segmentation models to quantify the carbides in the test set. As shown in Figure 7, the established models (DeepLabV3Plus and PSPNet) with the highest mIoU values outperformed other models in terms of quantitative accuracy. The established

PSPNet model achieved the lowest error with strong data augmentation and receptive field enlargement. The predicted total carbide number and average carbide radius differed from the actual values by 29 and 5.39 nm, respectively.

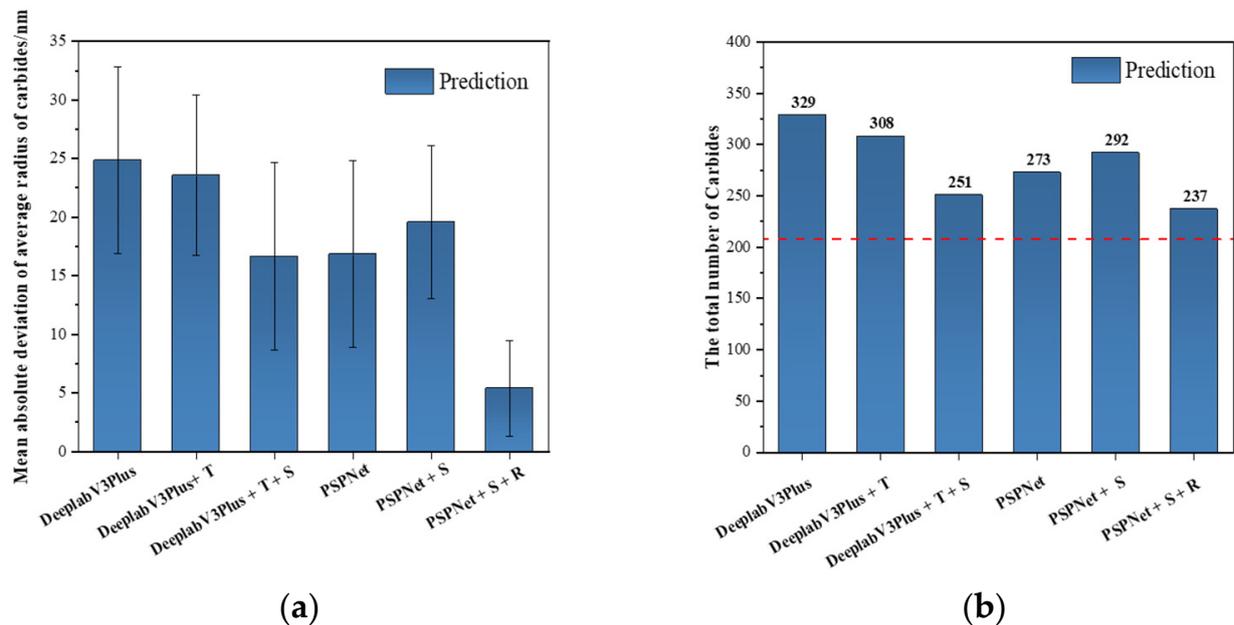


Figure 7. Quantification results of different models. (a) Quantification of average carbide radius. (b) Quantification of total carbide number. T, S, and R denote the use of transfer learning, strong data augmentation, and receptive field enlargement, respectively. The red line represents the true value.

To demonstrate the advantages of the deep learning method, we compared it to the traditional binary method. A representative image is shown in Figure 8a, and the ground truth labeling result is shown in Figure 8b. Using the OpenCV library, we calculated the actual carbide number (44) and average carbide radius (91.01 nm). According to the quantization of the above PSPNet model, as shown in Figure 8c, we get the predicted values of 51 and 79.27 nm, respectively. The predicted and actual values are highly consistent. Figure 8f,g shows the trends of the carbide number and average carbide radius with changes in the threshold. Figure 8d,e shows the binary plots corresponding to the carbide number closest to the real value and the average carbide radius closest to the real value, respectively. As shown in the figures, the relationship between the threshold and the quantitative results is essentially linear, and the optimal binarization accuracy is similar to that of the deep learning model. However, the thresholds used in Figure 8d,e are different and have a significant difference. Additionally, the binarized images have a large amount of noise, and the matrix or the carbide has been basically connected, losing the morphological information. The deep learning method completely avoids these problems and obtains the optimal quantification results through automatic model calculation without any human input.

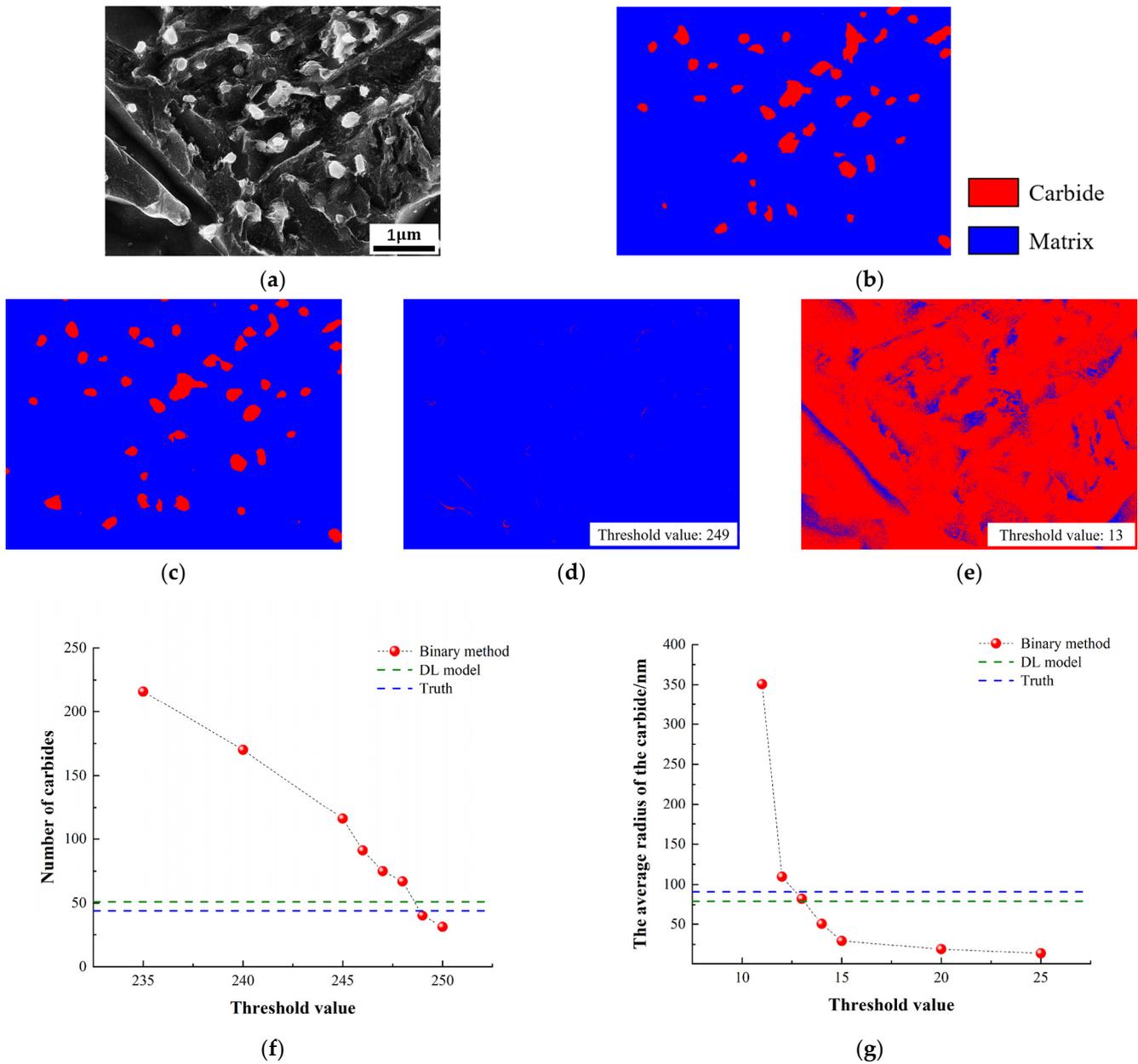


Figure 8. Comparison of the method in this study and the traditional binary image method: (a) selected SEM image; (b) labeled image; (c) deep learning segmentation image; (d,e) binary plots of the quantitative results closest to the actual carbide number and average carbide radius, respectively; (f,g) graphs of the variation of the carbide number and average carbide radius as the threshold changes.

5. Conclusions

In this work, we trained dozens of semantic segmentation models on the granular carbide dataset and the blocky ferrite dataset to study the impact of common training techniques on model accuracy. This study provides systematic guidance for the segmentation and quantitative analysis of steel material datasets. Our main findings are:

1. Transfer learning, strong data augmentation, and enlarging the receptive field can improve segmentation accuracy in most cases, improve the model's ability to microstructure segmentation, and reduce the area of misclassified regions.
2. Stacking multiple beneficial training techniques that improve segmentation accuracy leads to more accurate semantic segmentation models. Evaluation results demon-

strate a 1–2.5% increase in mIoU for DeepLabV3Plus and PSPNet models across both datasets.

3. To quantify the average radius and total number of the carbides of the test set, we applied optimal segmentation models. The established PSPNet model with strong data augmentation and receptive field enlargement achieved predicted deviations of 5.39 nm and 29 from the actual values, respectively. These results agree well with the ground truth. Additionally, the PSPNet model does not require manual input and generates more reasonable and accurate segmented images than the traditional binarization method.

Author Contributions: Conceptualization, X.M. and Y.Y.; methodology, X.M.; software, Y.Y.; validation, X.M. and Y.Y.; formal analysis, X.M.; investigation, X.M.; resources, X.M.; data curation, X.M.; writing original draft preparation, X.M.; writing review and editing, Y.Y.; visualization, X.M.; supervision, X.M.; project administration, Y.Y.; funding acquisition, Y.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key Research and Development Program of China (No. 2022YFB3707501), National Natural Science Foundation of China (No. 52211530455), Natural Science Foundation of Jiangsu Province (BK20230502) and the Jiangsu Funding Program for Excellent Postdoctoral Talent.

Data Availability Statement: The data presented in this paper are available on request from the corresponding author.

Acknowledgments: The authors are very grateful to the reviewers and editors for their valuable suggestions, which have helped improve the paper substantially.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. DeCost, B.L.; Lei, B.; Francis, T.; Holm, E.A. High Throughput Quantitative Metallography for Complex Microstructures Using Deep Learning: A Case Study in Ultrahigh Carbon Steel. *Microsc. Microanal. Off. J. Microsc. Soc. Am. Microbeam Anal. Soc. Microsc. Soc. Can.* **2019**, *25*, 21–29. [[CrossRef](#)] [[PubMed](#)]
2. Dehoff, R.; Russ, J. *Practical Stereology*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2001.
3. Martyushev, N.V.; Egorov, Y.P.; Utiev, M. Computer analysis of the material structure. In Proceedings of the 8th International Scientific and Practical Conference of Students, Post-Graduates and Young Scientists Modern Technique and Technologies, MTT 2002, Tomsk, Russia, 12 April 2002; pp. 159–161.
4. Martyushev, N.V.; Egorov, Y.P. Determination of the signal strength with the computer analysis of the material structure. In Proceedings of the 9th International Scientific and Practical Conference of Students, Post-Graduates Modern Techniques and Technologies, MTT 2003, Tomsk, Russia, 7–11 April 2003; pp. 192–194.
5. Stuckner, J.; Frei, K.; McCue, I.; Demkowicz, M.J.; Murayama, M. AQUAMI: An open source Python package and GUI for the automatic quantitative analysis of morphologically complex multiphase materials. *Comput. Mater. Sci.* **2017**, *139*, 320–329. [[CrossRef](#)]
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444. [[CrossRef](#)] [[PubMed](#)]
7. Shunmuga Perumal, P.; Wang, Y.; Sujasree, M.; Tulshain, S.; Bhutani, S.; Suriyah, M.K.; Kumar Raju, V.U. LaneScanNET: A deep-learning approach for simultaneous detection of obstacle-lane states for autonomous driving systems. *Expert Syst. Appl.* **2023**, *233*, 120970. [[CrossRef](#)]
8. Hoque, S.; Xu, S.; Maiti, A.; Wei, Y.; Arafat, M.Y. Deep learning for 6D pose estimation of objects—A case study for autonomous driving. *Expert Syst. Appl.* **2023**, *223*, 119838. [[CrossRef](#)]
9. Liang, G.; Zheng, L. A transfer learning method with deep residual network for pediatric pneumonia diagnosis. *Comput. Methods Programs Biomed.* **2020**, *187*, 104964. [[CrossRef](#)]
10. Lee, C.; Liao, Z.; Li, Y.; Lai, Q.; Guo, Y.; Huang, J.; Li, S.; Wang, Y.; Shi, R. Placental MRI segmentation based on multi-receptive field and mixed attention separation mechanism. *Comput. Methods Programs Biomed.* **2023**, *242*, 107699. [[CrossRef](#)]
11. Cui, W.; Zhang, Y.; Zhang, X.; Li, L.; Liou, F. Metal Additive Manufacturing Parts Inspection Using Convolutional Neural Network. *Appl. Sci.* **2020**, *10*, 545. [[CrossRef](#)]
12. Ma, B.; Ban, X.; Huang, H.-Y.; Chen, Y.; Liu, W.; Zhi, Y. Deep Learning-Based Image Segmentation for Al-La Alloy Microscopic Images. *Symmetry* **2018**, *10*, 107. [[CrossRef](#)]
13. Shen, C.; Wang, C.; Huang, M.; Xu, N.; van der Zwaag, S.; Xu, W. A generic high-throughput microstructure classification and quantification method for regular SEM images of complex steel microstructures combining EBSD labeling and deep learning. *J. Mater. Sci. Technol.* **2021**, *93*, 191–204. [[CrossRef](#)]

14. Breumier, S.; Martinez Ostormujof, T.; Frincu, B.; Gey, N.; Couturier, A.; Loukachenko, N.; Aba-perea, P.E.; Germain, L. Leveraging EBSD data by deep learning for bainite, ferrite and martensite segmentation. *Mater. Charact.* **2022**, *186*, 111805. [[CrossRef](#)]
15. Zhang, Q.-S.; Zhu, S.-C. Visual interpretability for deep learning: A survey. *Front. Inf. Technol. Electron. Eng.* **2018**, *19*, 27–39. [[CrossRef](#)]
16. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* **2021**, *109*, 43–76. [[CrossRef](#)]
17. Ma, J.; Hu, C.; Zhou, P.; Jin, F.; Wang, X.; Huang, H. Review of Image Augmentation Used in Deep Learning-Based Material Microscopic Image Segmentation. *Appl. Sci.* **2023**, *13*, 6478. [[CrossRef](#)]
18. Zhao, H.; Shi, J.; Qi, X.; Wang, X.; Jia, J. Pyramid Scene Parsing Network. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 6230–6239.
19. Chen, L.-C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; pp. 833–851.
20. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the 33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, BC, Canada, 8–14 December 2019.
21. Li, M.; Xie, X.; Zheng, M. OpenMMLab Semantic Segmentation Toolbox and Benchmark. Available online: <https://github.com/open-mmlab/msegmentation> (accessed on 1 October 2023).
22. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollar, P. Focal Loss for Dense Object Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *42*, 318–327. [[CrossRef](#)] [[PubMed](#)]
23. Stuckner, J.; Harder, B.; Smith, T.M. Microstructure segmentation with deep learning encoders pre-trained on a large microscopy dataset. *NPJ Comput. Mater.* **2022**, *8*, 200. [[CrossRef](#)]
24. Halevy, A.; Norvig, P.; Pereira, F. The Unreasonable Effectiveness of Data. *IEEE Intell. Syst.* **2009**, *24*, 8–12. [[CrossRef](#)]
25. Sun, C.; Shrivastava, A.; Singh, S.; Gupta, A. Revisiting Unreasonable Effectiveness of Data in Deep Learning Era. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 843–852.
26. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1345–1359. [[CrossRef](#)]
27. Yosinski, J.; Clune, J.; Bengio, Y.; Lipson, H. How transferable are features in deep neural networks? In Proceedings of the Advances in Neural Information Processing Systems (NIPS) 27: 28th Annual Conference on Neural Information Processing Systems 2014, Montreal, QC, Canada, 8–11 December 2014.
28. Feng, S.; Zhou, H.; Dong, H. Application of deep transfer learning to predicting crystal structures of inorganic substances. *Comput. Mater. Sci.* **2021**, *195*, 110476. [[CrossRef](#)]
29. Deng, J.; Dong, W.; Socher, R.; Li, L.J.; Kai, L.; Li, F.-F. ImageNet: A large-scale hierarchical image database. In Proceedings of the 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 20–25 June 2009; pp. 248–255.
30. Bartolini, I.; Moscato, V.; Postiglione, M.; Sperli, G.; Vignali, A. Data augmentation via context similarity: An application to biomedical Named Entity Recognition. *Inf. Syst.* **2023**, *119*, 102291. [[CrossRef](#)]
31. Devries, T.; Taylor, G.W.J.A. Improved Regularization of Convolutional Neural Networks with Cutout. *arXiv* **2017**, arXiv:1708.04552.
32. Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; Yang, Y. Random Erasing Data Augmentation. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 13001–13008. [[CrossRef](#)]
33. Kingma, D.P.; Welling, M.J.C. Auto-Encoding Variational Bayes. *arXiv* **2013**, arXiv:1312.6114.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.