A migration learning based multi-headed attentional convolutional neural network applied to metallographic image denoising

Shicheng Xie[1,2], Lixin Tang[1] and Yong Shuai[3]

[1] National Frontiers Science Center for Industrial Intelligence and Systems Optimization, Northeastern University,
Shenyang, 110819, China.
[2] Key Laboratory of Data Analytics and Optimization for Smart Industry (Northeastern University), Ministry of Education, Shenyang, 110819, China.
2110275@stu.neu.edu.cn
[3] Jiangxi Xinyu Iron and Steel Co.,Ltd, Xinyu, 338001, China.
shy@xinsteel.com.cn

Abstract. The rapid development of the manufacturing industry has increased the demand for metal materials, which has also put higher requirements on the quality of metal materials. As the most widely used metal, steel metallography is an important tool for steel quality assessment and material analysis. However, due to the presence of noise, metallographic images suffer from serious quality loss and information blurring, while the main problem in materials science and engineering is that the data set is too small to meet the deep learning modeling conditions. To solve this problem, we propose an end-to-end image denoising TransDnCNN model. In this study, the model is first pre-trained on a large-scale image dataset and migrated to a steel metallographic dataset using two migration learning methods: freezing all and fine-tuning some of the convolutional layer weights, while introducing a multi-headed attention mechanism to capture the relationship between different scales and attention points in the image to learn the noise distribution and texture feature representation of the image. The experimental results show that our proposed TransDnCNN has a significant performance improvement on the steel metallographic image denoising task. The successful application of the model provides a reliable image denoising solution for the steel industry and helps to improve the accuracy and efficiency of steel quality assessment and material analysis.

Keywords: Steel metallographic images, TransDnCNN, Migration learning,multi-headed attention mechanism,Image denoising.

1 Introduction

Metallographic examination is one of the simplest, widespread, and effective research and testing methods in the field of materials science and engineering. It is also a

powerful tool for analyzing and studying various materials, establishing microstructure and properties, studying structural transformation dynamics, etc. However, due to the noise and other interference factors in the acquisition process, metallographic images [1] usually present blurred and distorted conditions, which make the accurate analysis and interpretation of steel tissues and properties difficult. In order to improve the results while reducing the uncertainty caused by subjective and objective factors, experts and scholars have devoted themselves to solving these problems with computer vision automation methods [2].

Image denoising is an important task in the field of computer vision. The goal is to recover noise-contaminated images to make them clearer and more visual, and it is of great practical importance to develop efficient and accurate image denoising methods. In the past decades, many traditional machine learning methods have emerged in the field of image denoising. These methods are usually improved based on statistical models and mathematical filters, such as mean filtering [3], median filtering [4], wavelet transform [5], etc. Algorithms such as KSVD [6] and BM3D [7] model and exploit similar small patches based on the similarity of non-local texture information of image patches, and thus remove noise. Another common approach is to construct regular terms to constrain the relationship between the noise and the original image based on the image's prior knowledge to achieve the goal of noise removal. WNNM [8] represented the image as a linear combination of a set of sparse coefficients and a set of dictionary basis functions and achieved the effect of noise removal by minimizing the kernel norm of the sparse representation. Then Zhang et al. [9] constructed a Wiener filter and low-rank regularization based on the complementary local and global information of the image, and used the alternating direction multiplier method to solve the problem in order to achieve the goal of noise removal. Traditional image denoising methods based on regular terms usually use optimization algorithms to solve the optimal solution of the problem, for example, least squares-based methods [10], variational model-based methods [11], and so on. These methods can reduce noise and recover image details to some extent, but there are some limitations in dealing with complex noise and image structures. Meanwhile, traditional methods require manual setting of regularization parameters and trade-off terms, which require specialized design and adjustment for different images and noise types.

Unlike traditional methods, deep learning methods have the ability to automatically learn image feature representations using multilayer neural network models, which can extract high-level abstract features from large amounts of data and thus denoise images more accurately and have stronger generalization capabilities. The most representative one is the convolutional neural network (CNN), which can be regarded as a filter to optimize the network parameters by learning the mapping relationship from the noisy image to the denoised image using back propagation. Many deep learning algorithms are optimized at this stage, and ANN et al. [12] proposed a combined non-local self-similarity prior and convolutional neural network (CNN) denoising framework with a better denoising effect on complex texture images; Cruz C. et al. [13], on the other hand, proposed a CNN local multiscale noise reduction and a non-local filter (NLF) based fusion framework, NN3D, which also achieved good

denoising results. Also considering the local and non-local similarity of image textures and constructing the neighborhood dynamically in the feature space, Valsesia et al. [14] improved on the convolutional neural network architecture and proposed a neural network with graph-convolutional layers to accomplish the denoising task. Meanwhile, the probabilistic statistics-based image denoising technique GAN model [15] is able to learn the conditional probability distribution between noisy and clean images and thus achieve image denoising. However, CNN and GAN models are relatively weak in modeling global and non-local texture information, and they cannot provide accurate guidance for specific noise regions. To address this limitation, the Attention Mechanism (AM) was introduced to enhance the model's ability to perceive image features. The attention mechanism can automatically focus on the important features in an image, and Xu et al. [16] proposed a spatial channel attention mechanism based on the deep learning U-Net model, which in turn helps identify structures and evaluate their performance for effective segmentation of different classes. Zhang et al. [17], on the other hand, proposed the RatUNet model, which is a self-attentive mechanism in both space and channel to guide the CNN for image denoising, making the final denoised image smoother and sharper than other methods.

In this paper, we introduce the idea of migration learning into TransDnCNN, a steel metallographic denoising model with a multilayer convolutional neural network and multi-head attention mechanism, and take advantage of the attention mechanism to fully exploit the key feature information in the image to improve the denoising effect and recover more realistic image details.

The rest of the paper is organized as follows: Section 2 briefly describes the relevant theories used before modeling. The proposed DnCNN model architecture is given in Section 3, followed by the use of multi-headed attention, and finally the expressions of the loss function and the noise solution formulas are given. In Section 4, experiments are conducted to verify the effectiveness of the TransDnCNN model in the field of material metallography. Finally, a summary is given in Section 5.

2 Related works

In this section, several conceptual modules of deep learning image denoising are elaborated in order to introduce our proposed model.

2.1 Convolutional layer

The convolutional layer is the core of a convolutional neural network. Convolutional neural networks learn high-level features in the input samples by convolution. The convolution takes the input, uses a convolutional kernel, and outputs a feature mapping after computation by an activation function, whose input can be either the original data or another feature mapping of the convolutional output. As shown in Fig. 1, the convolution operation can be thought of as an inner product operation on the input data and the convolution kernel.

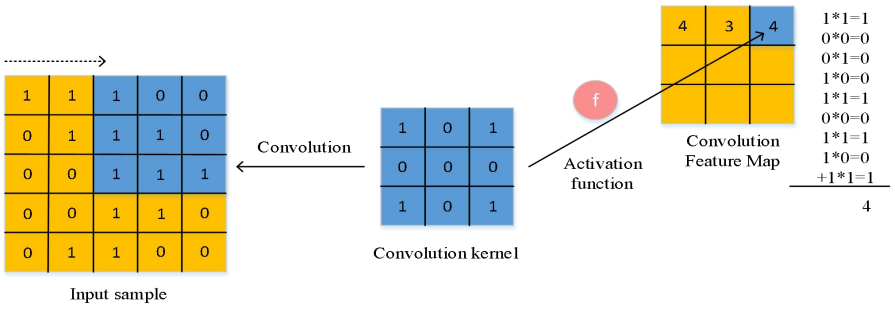


Fig. 1. Convolution operation

2.2 ReLU Function

The ReLU function, shown in Fig. 2 is a non-saturating activation function, that is also a non-linear function and has been a popular activation function for deep learning in recent years. The mathematical expression of the ReLU function is shown below:

$$f(u) = \max(u, 0) \tag{1}$$

When the ReLU function is used as the activation function, the output is equal to the input when the input u is greater than 0. It will not saturate, which can make the network converge faster and also solve the gradient disappearance problem; meanwhile, the ReLU function is simple to compute, which can improve the network efficiency.

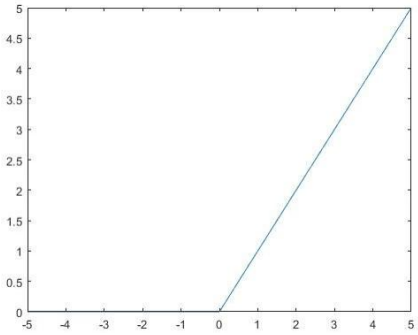


Fig. 2. Geometric image of the ReLU function

2.3 Transfer learning

Transfer learning is used to solve the problem of modeling small sample datasets. For the problem that metallographic image datasets are often small and cannot satisfy deep learning network modeling, transfer learning can accelerate the learning process of the target task and improve the generalization ability of the model with rich data and knowledge of existing tasks. The essence of migration learning is that the model developed for Task A is used as the base model and reused in the process of

developing the model for Task B. The migration learning method is based on the following principles: feature-based migration, instance-based migration, and shared parameter-based migration. The migration learning methods are feature-based migration, instance-based migration, and shared parameter-based migration.

3 Methodology

3.1 The overall architecture of the network

The network as a whole is divided into two main parts. Firstly, a convolution filter of size 3×3 is set in the first layer, and because it is a grayscale image, it is set to generate 64 feature maps of $3\times3\times1$. Then a linear correction is performed using the nonlinear Relu activation function. In layers 2-16, we refer to the convolution, normalization, and activation function operations collectively as a module, here using 64 filters of size $3\times3\times64$. In layer 17, a single image is generated using a filter of size $3\times3\times64$. The above is the first part. Subsequently, the image is divided into n small patches of size 40×40 according to the step size of 1 and then entered into the norm layer for linear transformation to form the matrix $n\times1600$. The forward propagation dimension is set to 2048, and the dropout parameter is 0.1. The MLP layer receives the input attention weights, maps and transforms the attention weights nonlinearly, reconverts them into n image patches of size 40×40, and links the residuals twice after the attention layer and the MLP layer. The above is the second part of the overall architecture.

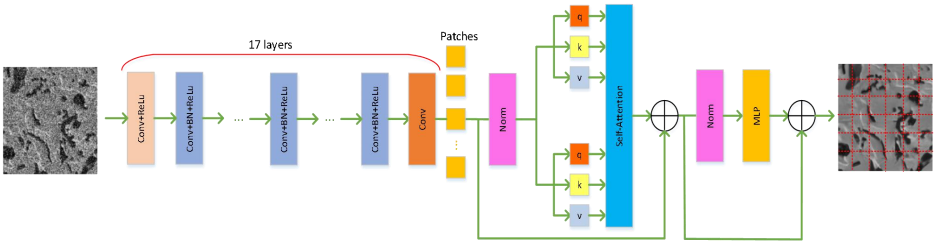


Fig. 3. Overall architecture of the TransDnCNN model with multi-headed attention

3.2 Multi-headed attention mechanism module

The Transformer [18] proposed by the Google team was first successful in natural language processing tasks, and then inspired by NLP, the VIT [19] model, based on the combination of convolutional neural networks (CNNs) and self-attention, was proposed to handle a wide range of tasks in computer vision. Transformer model is further divided into encoder and decoder, and in this paper we only access the encoder part after layer operation. The self-attention mechanism can capture and learn the long-distance non-local spatial feature information of the metallographic map, more fully exploit the texture features of the metallographic map, and solve problems such as gradient disappearance and explosion. Each head in the attention head module can learn different attention weights to capture the correlation between the input images,

as shown in Fig. 3. Above, the image is divided into several small patches after a multiple layer operation, and each of them is normalized and linearly transformed, followed by a multi-head attention operation with the following Eq. 2:

$$Attention(Q,K,V)=\text{Softmax}(\frac{QK^T}{\sqrt{d_k}})V \tag{2}$$

where d denotes the number of channel dimensions of the input image, while ensuring that the dimensions of $Q(Query)\in R^n$, $K(Key)\in R^{n\times d_k}$, and $V(value)\in R^{n\times d_v}$ are consistent. The input Q is point multiplied with K^T , respectively, to calculate the attentional similarity and obtain the weights; $\sqrt{d_k}$ denotes the scaling factor, which is used to balance the gradient size in the attentional calculation, and is multiplied with V . Finally, the weights are normalized using the Softmax function. Repeating this operation for an attention mechanism with n heads can be expressed as:

$$\begin{aligned} &MultiHead(Q,K,V)=Concat(head_1,\ldots head_n)W \\ &where \quad head_i=Attention(QW_i^Q,KW_i^k,VW_i^V) \end{aligned} \tag{3}$$

In the above equation, $W_i^Q\in R^{d_{model}\cdot d_k}, W_i^K\in R^{d_{model}\cdot d_k}, W_i^V\in R^{d_{model}\cdot d_k}$ is the parameter matrix of the linear mapping, and $W\in R^{d_{model}\cdot d_k}$ denotes the weight matrix. In order to maintain the information transfer of the metallographic features and prevent the gradient from disappearing, the outputs of the n attention heads are linearly transformed as well as weighted and summed, and the residual connections are added so that the original input sequence is summed with the output of the attention module. The use of a multi-headed attention mechanism can provide richer information about the texture features of metallographic maps from different spaces.

3.3 Image denoising network loss function and residual learning

The image denoising task itself can actually be reduced to $y=x+v$,whether it is image denoising using traditional convolutional neural networks [20] or SVD image decomposition denoising algorithms based on machine learning algorithms [21], which are trained and learned to map function $F(y)=x$ to obtain a clean image closest to the original image. The TransDnCNN loss function proposed in this paper can be expressed by the following equation:

$$l(\Theta)=\frac{1}{2N}\sum_{i=1}^{N}\left\|\Re(y_i;\Theta)-(y_i-x_i)\right\|_F^2 \tag{4}$$

In the above equation, Θ is the trainable parameter in the model, $\{(y_i - x_i)\}_{i=1}^{N}$ denotes the difference between the N noisy images and the sample image matching patches, and $\Re(y_i;\Theta)$ denotes the result map coming from the output of the model. The smaller the difference between the two, the better result obtained. The introduction of the Transformer module with residual connections allows for better learning of the mapping relationships, obtaining smaller loss values while achieving faster convergence, and using a normalization approach to resolve the internal parameter shifts of the network parameters [22], which in turn allows the above equation to be rewritten as:

$$l(\Theta) = \min_{x} \Psi(y-x) + \lambda_1 \sum_{k=1}^{K}\sum_{p=1}^{N} \rho_k \left(\left(f_k * x \right)_p \right) + \lambda_2 \left(\frac{1}{n} \sum KL(P_{(i)} \log(\frac{P_{(i)}}{Q_{(i)}})) \right) \quad (5)$$

λ_1, λ_2 is the regularization parameter, N denotes the input image patch size, $f_k * x$ denotes the current image patch with the k-th convolution kernel of size f for convolution operation, and $\rho_k(\bullet)$ denotes the k-th adjustable penalty function, which is taken as the first regularization term. P is the target distribution, Q is the distribution predicted by the model, P(i) and Q(i) denote the attention weights of P and Q at position i, respectively, and the KL scatter is used to measure the difference between n attention distributions and as the attention loss function, i.e., the second canonical term.

In the experiments of this paper, Gaussian white noise is added to the original image; let it be $\Psi(z) = \frac{1}{2}\|z\|^2$. Subsequently, the stochastic gradient descent algorithm (SGD) with momentum optimization or the Adam [23] algorithm is used to solve it, giving the diffusive iterative gradient descent expression at pixel point y:

$$x_1 = y - \alpha\lambda_1 \sum_{k=1}^{K} \left(\bar{f}_k * \phi_k \left(f_k * y \right) \right) - \lambda_2 \left(\frac{1}{n} KL(\bar{P} \| \bar{Q}) \right) - \alpha \left. \frac{\partial \Psi(z)}{\partial z} \right|_{z=0} \quad (6)$$

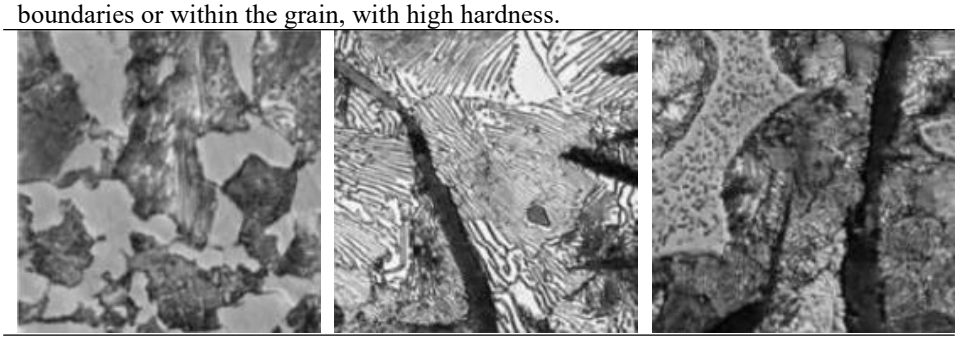
α denotes the step of descent, ϕ_k is the derivative of ρ_k, $\bar{f}_k$ is the accompanying filter, $\bar{P}, \bar{Q}$ denotes the updated target and predicted distribution at different rounds, and the horizontal and vertical smoothing operation for a single pixel point z in order to minimize noise is denoted as $\left. \frac{\partial \Psi(z)}{\partial z} \right|_{z=0} = 0$. The residual formula is continuously used to learn A, which is combined with batch processing and normalization to accelerate the training and thus improve the denoising performance. It is worth noting that the model is applicable to many types of noisy distributions and is generalizable.

4 Experimental Results

4.1 Dataset description

For the training images, due to the difficulty and limited amount of metallographic data production, it cannot meet the modeling training requirements of TransDnCNN, so we use 400 Set12 datasets of size 180×180 for migration learning training, i.e., extracting features of image samples on small sample datasets and migrating the learned parameters and weights to the TransDnCNN model denoising task.

For the test images, in order to fully verify the effectiveness of the TransDnCNN model proposed in this paper, we used several important phases of steel with different carbon contents and evaluated them. All the steel metallographic samples were obtained by flat cutting with a metallographic cutting machine, and their surfaces were ground and polished; then the surfaces were wiped with a 96% alcohol + 4% nitric acid etching solution, and finally observed with a metallographic microscope at a magnification of 400 times. It should be noted that all kinds of metallographic sample images are essentially RGB, three-channel color images. In order to ensure that the features do not decay while simplifying the image denoising process, all color images are converted into grayscale images in this paper. Fig. 4 (a)–(l) below lists 12 representative metallographic microstructures and describes the crystal structure of the alloys within. Among them austenite: Steel is heated to a high temperature so that its crystal structure changes to austenite. Ferrite: Austenite undergoes a relatively slow cooling process, which transforms it into ferrite. This process is called austenite transformation or annealing. Ferrite is usually formed at lower temperatures, and it has a more stable crystal structure than austenite. Martensite: A highly hard metallic crystal structure with high hardness and brittleness formed through the transformation of austenite in rapid cooling, usually occurring during the cold rolling and quenching of steelmaking. Pearlitic: Composed of two distinct phases, austenite and ferrite, this structure is formed in mild steel during cooling over a range of temperatures and has a certain degree of hardness and strength. Bainite: Formed in steel by moderate cooling rates at moderate temperatures, the formation temperature is usually between that of martensite and pearlite. Carburite: Formed during the quenching and tempering process of steelmaking, where carbon is precipitated in the form of carbides at grain boundaries or within the grain, with high hardness.



(a)Reticulated ferrite + Pearlite | (b)Flake graphite + pearlescent | (c)Flake graphite + Pearlite + Phosphorus eutectic

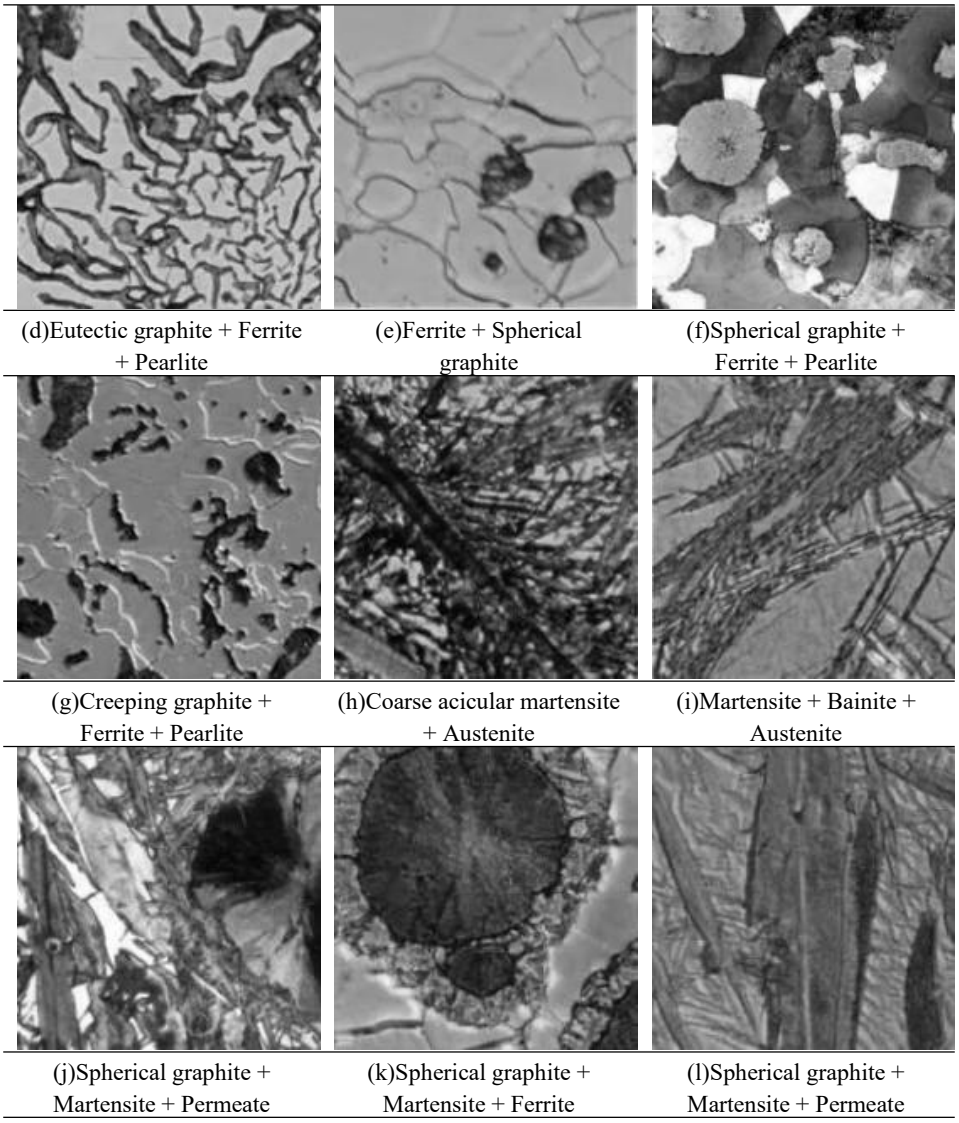


Fig. 4. Twelve metallographs with different characteristics at different temperatures and carbon contents were used for the test set.

4.2 Parameters settings

(1)**Training set and test set image preprocessing operations**: In the training set, in order to increase the number of samples, we crop the metallographic map to size 40×40 image patches and get a total of 1600×128 small patches to train the model. In the test set, to ensure the same dimensionality as the Transformer module, Fig. 4 is the original RGB metallographic image after grayscale transformation, where (a), (f)

is the intercepted upper left corner; (i) is the intercepted lower left corner; (b), (e), (h) is the intercepted lower right corner; (c)-(d), (g), (j)-(l) are obtained by intercepting the middle part of the whole image, all of size 180×180 .

(2)**Parameter settings:** We add Gaussian white noise of $\sigma\in[0,55]$ to the test set images, set the network depth of TransDnCNN to 17 layers, set the number of attention heads to 2 in order to verify the feasibility of the model while improving the operational efficiency, set the transformer-layer to 6, initialize the weights using the method in the literature [24], and use SGD to update the parameters. The momentum is set to 0.9, the weight decay is 0.0001, the batch is set to 32, and the learning rate decays from 1e-1 to 1e-4 in order to make the model converge in effective time with 30 training epochs.

(3)**Experimental environment:** The algorithms in this paper are performed on a 64-bit Windows 10 platform with a hardware environment of an Intel (R) Core (TM) CPU i7-8750H at 2.21 GHz and 16 GB of RAM. TransDnCNN uses Python 3.6, the Pytorch 1.10.2 platform framework, and Matlab R2019b for experimental simulation. It takes about 12 hours to train the model using a single GPU, the NVIDIA GTX1060.

4.3 Numerical Statistics and Visualization

In order to evaluate the denoising effect of the proposed model, the peak signal-to-noise ratio (PSNR) [25] is induced. The equation of:

$$PSNR=10\lg\frac{255\times255}{\frac{1}{mn}\sum_{i=1}^{M}\sum_{j=1}^{N}[x(i,j)-\hat{x}(i,j)]^2} \tag{7}$$

Where $x(i,j)$ denotes the original sample image, $\hat{x}(i,j)$ denotes the denoised image, and m and n denote the height and width of the image. A higher PSNR value means that the denoised image is closer to the original image. Table 1 below shows the PSNR values of the 12 test set images at Gaussian noise factor $\sigma=15,25,50$. Fig. 5 indicates the visualization of the noise map and output map under three different noise levels as a group of four different kinds of metallographic maps of the same size (40×40).

Table 1. PSNR values of 12 images after denoising at different σ levels

PSNR /dB	(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)	(i)	(j)	(k)	(l)
$\sigma=15$	24.59	24.51	24.42	24.67	24.54	24.61	24.65	24.79	24.30	24.20	24.75	24.55
$\sigma=25$	20.23	20.68	20.21	20.15	19.96	20.37	20.05	20.49	20.21	20.38	19.98	20.30
$\sigma=50$	14.46	15.20	14.43	14.43	14.62	14.69	14.55	15.02	14.41	14.78	14.62	14.73

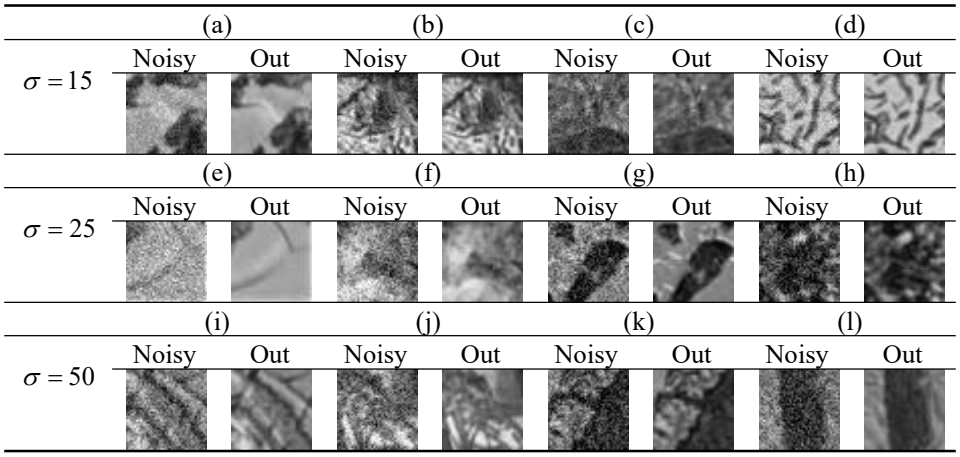


Fig. 5. Noise and denoised images of 12 images at different σ levels (size 40×40)

4.4 Discussions and analysis

Through Table 1, it can be found that after the test set images are denoised by the TransDnCNN model, the PSNR values are maintained at a certain level at $\sigma = 15,25,50$ time, and with the increase of the noise parameter values, the PSNR values become a decreasing trend, which also verifies the universality and reasonableness of the model, and the a priori expert information can be further queried to fine-tune the internal parameters, thus higher quality denoised metallographic maps can be obtained. The future main work simultaneously adjusts the number of convolutional layers and the Transformer module to the number of encoder patches and the number of heads inside, which can be used to evolve a suitable network architecture by combining the idea of neural network architecture search (NAS) with unsupervised learning.

5 Conclusion

In this paper, we propose a multi-headed attentional convolutional neural network (TransDnCNN) model based on migration learning that first uses a multilayer convolutional neural network as a skeleton to fully exploit the key features of the metallographic map. Subsequently, a multi-headed attention mechanism is added to enhance the model's ability to perceive local and non-local key features of the image and to provide targeted guidance for the denoising model to better preserve image details and textures. Finally, we show the PSNR and visualization results of various categories of metallographic images, which visually show the smoothness and clarity of the images after denoising by the model, and experimentally prove the strong robustness and universality of the model under different noise factors and in different categories of metallograms. Overall, the TransDnCNN model provides an innovative

solution to the steel metallographic image denoising problem and provides a useful reference for research and application in related fields.

Acknowledgement. This work was supported by the Major Program of National Natural Science Foundation of China (72192830, 72192831), the 111 Project (B16009).

References

1. Cheng D., Sha W., Xu Z., et al.: Computer Vision Analysis on Material Characterization Images. Advanced Intel. Syst. 4(3), 2100158 (2022).
2. Kesireddy A., McCaslin S.: Application of image processing techniques to the identification of phases in steel metallographic specimens. In: New Trends in Networking, Computing, E-learning, Systems Sciences, and Engineering, pp. 425-430. Springer International Publishing, USA (2015).
3. Thanh D. N. H., Engínoğlu S.: An iterative mean filter for image denoising. IEEE Access 7, 167847-167859 (2019).
4. Erkan U., Enginoğlu S., Thanh D. N. H., et al.: Adaptive frequency median filter for the salt and pepper denoising problem. IET Image Proces. 14(7), 1291-1302 (2020).
5. You N., Han L., Zhu D., et al.: Research on image denoising in edge detection based on wavelet transform. Applied Sciences 13(3), 1837 (2023).
6. Scetbon M., Elad M., Milanfar P.: Deep k-svd denoising. IEEE Trans. on Image Proces. 30, 5944-5955 (2021).
7. Dabov K., Foi A., Katkovnik V., Egiazarian K.: Image denoising by sparse 3-D transform-domain collaborative fifiltering. IEEE Trans. on Image Proces. 16(8), 2080-2095 (2007).
8. Chen Y., Pock T.: Trainable nonlinear reaction diffusion: A flflexible framework for fast and effective image restoration. IEEE trans. on Pat. Analys. and Machine Intel. 39(6), 1256-1272 (2016).
9. Zhang Y., Kang R., Peng X., et al.: Image denoising via structure-constrained low-rank approximation. Neural Comput. and Applic. 32, 12575-12590 (2020).
10. Hirakawa K., Parks T. W.: Image denoising using total least squares. IEEE Trans. on Image Proces. 15(9), 2730-2742 (2006).
11. Hsieh P. W., Shao P. C., Yang S. Y.: A regularization model with adaptive diffusivity for variational image denoising. Signal Proces. 149, 214-228 (2018).
12. Ahn B., Cho N. I.: Block-matching convolutional neural network for image denoising. arXiv preprint arXiv:1704.00524 (2017).
13. Cruz C., Foi A., Katkovnik V., et al.: Nonlocality-reinforced convolutional neural networks for image denoising. IEEE Signal Proces. Letters 25(8), 1216-1220 (2018).
14. Valsesia D., Fracastoro G., Magli E.: Image denoising with graph-convolutional neural networks. In: 2019 IEEE international conference on image processing (ICIP), pp. 2399-2403. IEEE, Taipei (2019).
15. Singh P., Pradhan G.: A new ECG denoising framework using generative adversarial network. IEEE/ACM Trans. on Comput. Biology and Bioin. 18(2), 759-764 (2020).

16. Xu Y., Zhang Y., Zhang M., et al.: Quantitative Analysis of Metallographic Image Using Attention-Aware Deep Neural Networks. Sensors 21(1), 43 (2020).
17. Zhang H., Lian Q., Zhao J., et al.: RatUNet: residual U-Net based on attention mechanism for image denoising. PeerJ Computer Science 8, e970 (2022).
18. Vaswani A., Shazeer N., Parmar N., et al.: Attention is all you need. Advan. in neural inform. Proces. Sys. 30 (2017).
19. Dosovitskiy A., Beyer L., Kolesnikov A., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020).
20. Jain V., Seung S.: Natural image denoising with convolutional networks. Advan. in neural inform. Proces. Sys. 21, (2008).
21. Guo Q., Zhang C., Zhang Y., et al.: An efficient SVD-based method for image denoising. IEEE trans. on Circ. and Sys. for Video Tech. 26(5), 868-880 (2015).
22. Ioffe S., Szegedy C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: International Conference on Machine Learning, pp. 448–456. Pmlr, LILLE (2015).
23. Kingma D. P., Ba J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014).
24. He K.., Zhang X., Ren S., et al.: Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: IEEE Inter. Confere. on Comput. Vision, pp. 1026– 1034, IEEE, Santiago (2015).
25. Huynh-Thu Q., Ghanbari M.: Scope of validity of PSNR in image video quality assessment. Electronics letters 44, 800-801 (2008).