AI

MDPI

Article

Deep Learning for Super-Resolution in a Field Emission Scanning Electron Microscope

Zehua Gao [1], Wei Ma [1], Sijiang Huang [1], Peiyao Hua [1] and Chuwen Lan [1,2,*]

[1] School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China; gaozehua@bupt.edu.cn (Z.G.); mawei@bupt.edu.cn (W.M.); huangsij@bupt.edu.cn (S.H.); huapeiyao@bupt.edu.cn (P.H.)

[2] State Key Laboratory of Millimeter Waves, Southeast University, Nanjing 210096, China

* Correspondence: lanchuwen@bupt.edu.cn

Received: 7 September 2019; Accepted: 30 September 2019; Published: 15 October 2019

Abstract: A field emission scanning electron microscope (FESEM) is a complex scanning electron microscope with ultra-high-resolution image scanning, instant printing, and output storage capabilities. FESEMs have been widely used in fields such as materials science, biology, and medical science. However, owing to the balance between resolution and field of view (FOV), when locating a target using an FESEM, it is difficult to view specific details in an image with a large FOV and high resolution simultaneously. This paper presents a deep neural network to realize super-resolution of an FESEM image. This technology can effectively improve the resolution of the acquired image without changing the physical structure of the FESEM, thus resolving the constraint problem between the resolution and FOV. Experimental results show that the apply of a deep neural network only requires a single image acquired by an FESEM to be the input. A higher resolution image with a large FOV and excellent noise reduction is obtained within a short period of time. To verify the effect of the model numerically, we evaluated the image quality by using the peak signal-to-noise ratio value and structural similarity index value, which can reach 26.88 dB and 0.7740, respectively. We believe that this technology will improve the quality of FESEM imaging and be of significance in various application fields.

Keywords: deep learning; super-resolution; convolutional neural network; field emission scanning electron microscope; field of view

1. Introduction

A field emission scanning electron microscope (FESEM) is a scanning electron microscope that uses a field emission electron gun to generate electrons that converge into a very fine electron beam to irradiate the surface of the prepared sample. When the beam strikes the sample surface, it interacts with the sample and excites secondary electrons. The detection system collects a secondary electronic signal and converts it into a video signal based on a certain rule. After being amplified, the signal is sent to the picture tube of the synchronous scan, modulated, and then imaged.

The most important feature of an FESEM is its ability to scan ultra-high-resolution images, particularly with the latest digital image processing technology, which can provide high-magnification and high-resolution (HR) scanned images, and print or save the output instantly. These properties make an FESEM one of the most effective instruments to observe and analyze microscopic morphology, organization, and composition. This is why an FESEM is widely used in various fields, such as scanning electron microscopy for carbonate sediments and the imaging of bacteria in rocks [1], semiconductor superlattice imaging [2], and the quality prediction of noisome from maltodextrin-based proniosomes [3].

However, there is a problem with the FESEM when it comes to using it for locating a certain area of interest in the sample. Owing to the balance between the resolution and field of view (FOV), images with a large FOV have a relatively low resolution, which renders the details ambiguous, making an accurate location difficult to determine. By contrast, if we increase the resolution to such an extent that we can achieve an accurate location, the FOV becomes so limited that we cannot cover all possible target areas. This problem has largely restricted the application of the FESEM.

Image stitching [4] and interpolation methods have been traditionally used to improve the image resolution. However, image stitching requires high mechanical precision, and multiple imaging limits the imaging speed. An interpolation method for improving the image resolution identifies a specific pixel point in the actual captured image to calculate the logical pixel points around it and consider them to be supplementary pixels, thereby enhancing the overall resolution of the image. Common algorithms include the nearest neighbor interpolation [5], bilinear interpolation [6], and bicubic interpolation [7]. Although an interpolation method is fast and simple to implement, there are still defects that occur at the image boundaries, such as jagged ridges and blurring effects. Moreover, these interpolation methods have a common problem. The same algorithms are used for super-resolution tasks of different devices, which means that generated HR images do not necessarily have the best resemblance to the original HR images for a specific device.

Deep learning [8] is an active fields of machine learning, and its concept comes from artificial neural networks. There are billions of neurons in the human brain, and because of this, the human brain has comparable data processing capability to computers. The artificial neural network establishes a kind of neural network computing model by abstracting the human brain neurons and forming a large topology between the different neurons. Deep learning is to achieve complex data processing functions by stacking multi-layers artificial neural networks to form a deep neural network structure.

The deep convolutional neural network, as a typical structure of a deep neural network, has been widely applied in various types of supervised and unsupervised learning, such as image classification [9], style transfer [10], voice recognition [11], and natural language processing [12]. However, it was found that, as the depth of the neural network increases, the accuracy of a convolutional neural network has a tendency toward saturation or even degradation. In 2015, He et al. [13] proposed a new type of deep neural network, called the residual network. By fitting a residual function instead of a primitive function, they enabled a deep neural network to learn more efficiently, leading to a better performance. In recent years, the residual network has been used to solve all kinds of imaging problems, such as tomography [14], infrared polarization imaging [15], and magnetic resonance imaging [16].

This study proposes a novel deep neural network based on the residual network, which can effectively improve the resolution and quality of an image acquired by an FESEM. Our deep neural network considers the image acquired by the FESEM to be the input, and rapidly generates an HR image, while simultaneously obtaining a large FOV and achieving the noise reduction effect.

2. Method

Samples: Butterfly specimens (Figure 1) were chosen as the experimental sample from which the image datasets were obtained. The super-resolution task not only focuses on improving the resolution of the entire image, but also on improving the details of different textures or patterns. Butterfly wings are covered by flat scaly hair, and different varieties of butterflies have different scaly hair patterns, as shown in Figure 2, which makes them very suitable for use in image data. We clipped the wings of different butterfly specimens and stuck them on a carrier. Owing to the poor electrical conductivity of butterfly wings, it was necessary to use an instrument to coat the surface of the sample with a silver conductive film in order to avoid the charging effect.

Figure 1. Butterfly specimens.

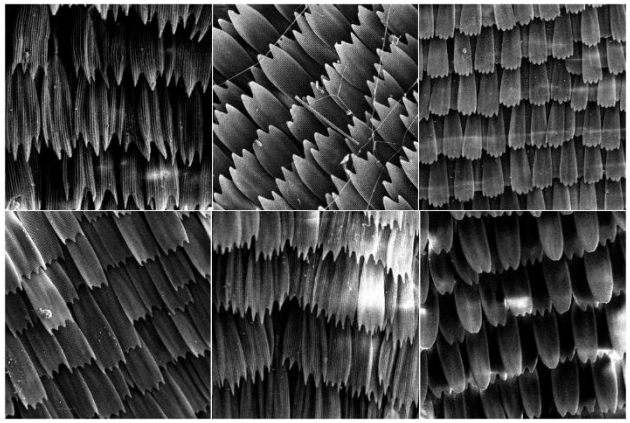

Figure 2. Different scaly hair patterns.

Image shooting: Images were acquired using a ZEISS Sigma 500 FESEM (Carl Zeiss AG, Jena, Germany), which adopts a mature GEMINI optical system design, and has a resolution of over 0.8 nm. A total of 1000 sets of image data were obtained, each of which contained a low-resolution (LR) image (1000× magnification) and an HR label image (2000× magnification).

Dataset: To obtain optimal results, the images used to train the neural network must be paired, i.e., LR input images with HR label images. As LR images have a large FOV, the HR images have a small FOV, and the image sets used in the super-resolution tasks need to have a constant FOV, the LR images were cropped to match the FOV of the HR images. Due to the positional shift when switching the magnification during shooting, we cannot directly crop an LR image at the center position. Pixel matching is required before we crop the LR images. This study uses the pixel matching method to reduce the HR image fourfold, and then scans the pixel to the best matching position in the LR image and crops it. In addition, considering that the dataset is relatively small, we used data expansion while creating our dataset. Original images were flipped horizontally and vertically, both individually and simultaneously. Finally, a total of 4000 sets of data were used in the training, validation, and testing of our deep neural network, where each set of images included an LR image (200 × 200) and an HR image (400 × 400).

3. Super-Resolution Deep Neural Network Structure

The structure of the super-resolution deep neural network is shown in Figure 3. Assume that the size of the input LR image is W × H × 3, where W and H represent the width and height of the image, and 3 represents three color feature maps, namely, red, green, and blue (RGB). First, the input image is applied with a mean subtraction conversion. The average value is subtracted from each independent

feature map, and the input data are normalized in each dimension, which not only avoids unnecessary numerical problems, but also enables the network to converge more quickly.

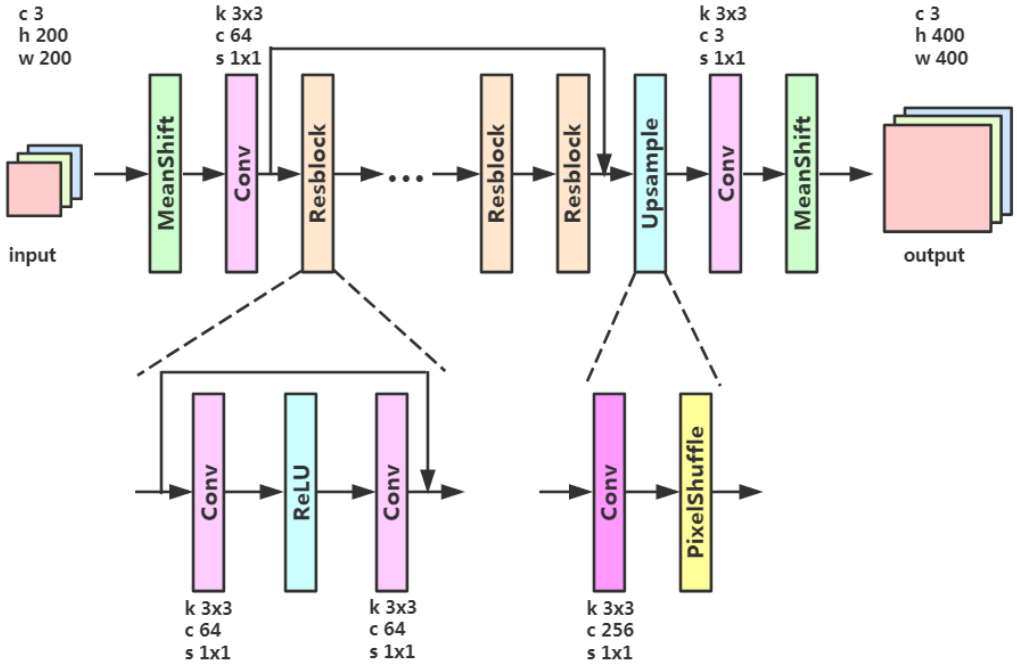

Figure 3. Super-resolution structure of the deep neural network.

The input convolutional layer maps 3 feature maps to 64 feature maps, as depicted in Figure 4. While adjusting the hyperparameters, the number of convolution kernels C was empirically set to 64, the convolution kernel size K was set to 3, the step size S was set to 1, and the zero padding P was calculated using P = K//2 (where '//' indicates a floor division). Thus, the size of the output image of the convolutional layer can be calculated according to the following formula:

$$W' = (W - K + 2P)/S + 1 \quad (1)$$

$$H' = (H - K + 2P)/S + 1 \quad (2)$$

$$C' = C = 64 \quad (3)$$

where W', H' and C' represent the width, height, and number of feature maps of the output image of the convolutional layer, respectively.

The convolutional layer is followed by 32 residual modules, each of which consist of two convolutional layers and an activation function. The convolutional layer hyperparameter is consistent with the input convolutional layer, and the activation function introduces nonlinear characteristics into the network, which enables the network to learn complex function mappings from the data. This paper uses the Rectified Linear Unit (ReLU) activation function, which is expressed as ReLU(x) = max(0, x). Thus, the calculation for each convolution module is as follows:

$$X_{n+1} = X_n + \mathrm{ReLU}(X_n \times W_n^{(1)}) \times W_n^{(2)} \quad (4)$$

where * denotes a convolution operation, X_n and X_{n+1} represent the input and output of the nth residual module, and $W_n^{(1)}$ and $W_n^{(2)}$ represent the parameter matrix of the first and second convolutional layers, respectively.

This is followed by an upsampling layer, which has a convolutional layer and a pixel conversion operation. The number of output feature maps of the convolutional layer is set to 4 times the number

of input feature maps. For example, if the input size of the convolution layer is W × H × 64, the output size after the convolution becomes W × H × 256.The pixel conversion operation is responsible for rearranging all feature map pixels, and the image size is changed from W × H × 256 to 2W × 2H × 64. Finally, after the convolutional layer and mean subtraction conversion, the number of image feature maps is changed from 64 to 3 (RGB).

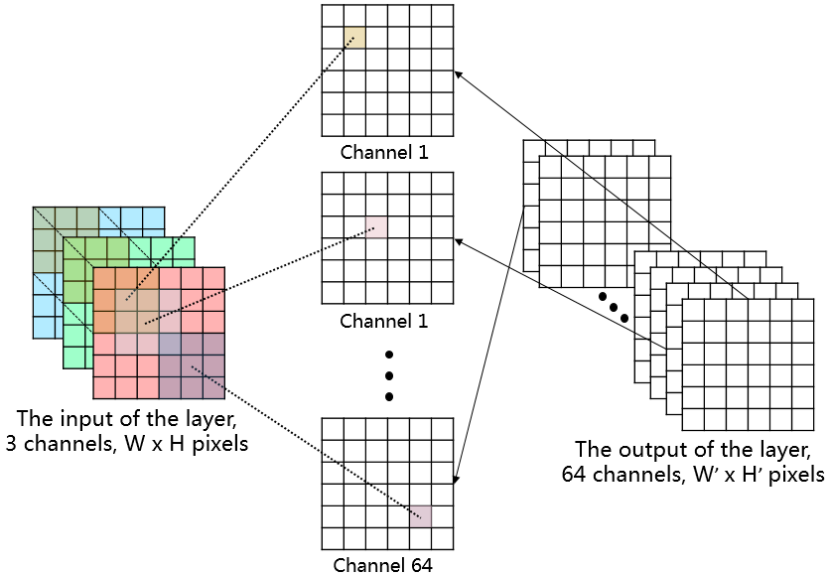

Figure 4. Details of the input convolutional layers.

4. Results

The dataset used to train the super-resolution deep neural network is obtained from the images of the butterfly wing specimen taken using an FESEM. To obtain optimum results while training, we enlarged the sample by 1000× and 2000× so as to obtain lower-resolution images of 200 × 200 pixels and higher resolution images of 400 × 400 pixels. There is a total of 4000 sets of image data, from which 3000 sets were randomly selected to be the training set for the neural network, and the remaining 1000 sets were used to verify the trained neural network to avoid over-fitting.

A schematic diagram of the training process of the super-resolution deep neural network is shown in Figure 5. An LR image is the input required for the neural network to generate an output image. This output image is compared with the HR label image to generate a loss function. The back-propagation algorithm [17] is employed to adjust the parameters in the super-resolution neural network to minimize the loss function. The final optimized neural network model is acquired through many training iterations.

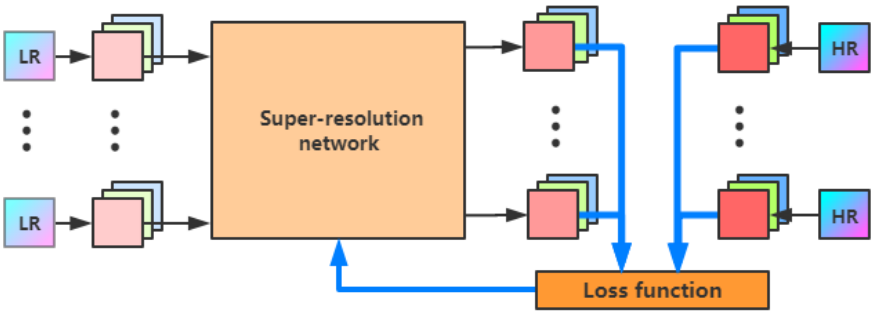

Figure 5. Training process of the deep neural network.

We use the L1 loss function [18] to optimize the network. The expression of the L1 loss function is as follows:

$$\text{loss}(X, Y) = \sum_{(c)} \sum_{(w)} \sum_{(h)} |Y_{c,w,h} - X_{c,w,h}| \quad (5)$$

where $Y_{c,w,h}$ and $Y_{c,w,h}$ represent the pixels at position (w, h) in the c^{th} feature map of the network output image and the HR label image, respectively. In addition, 2W and 2H indicate the width and height of the network output image, respectively. The errors of the network output image and the HR label image are calculated using the L1 loss function, and then propagated back to the network. The parameters in the network are optimized using the Adam optimizer, and the initial value of the learning rate is empirically set to 10^{-4}.

We trained a total of 300 epochs. After the completion of each training epoch with the training set, we validated the L1 loss using the validation set. The change in loss is shown in Figure 6. The blue line in the figure represents the training set loss, and the orange line represents the validation set loss. It can be seen from the validation set loss that the value of the loss has monotonically decreased, finally becoming stable. It means that the training of the network has not been over-fitted.

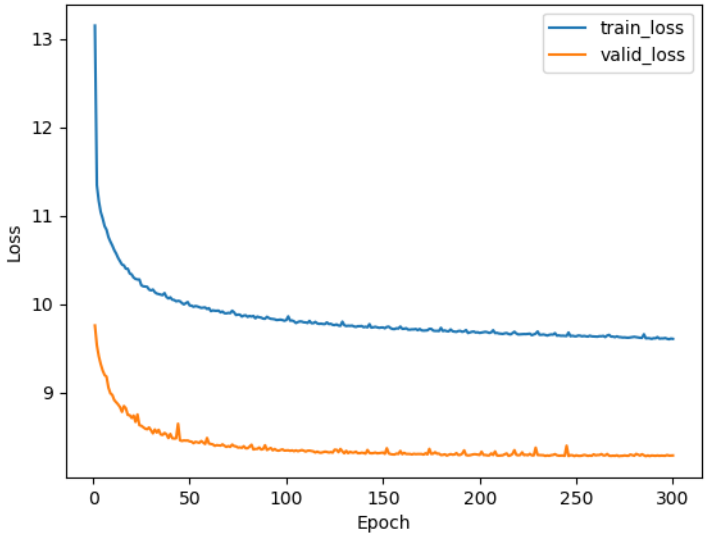

Figure 6. Training set loss curve and validation set loss curve.

At the time of verification, we also compared the network output image with the HR label image by calculating the peak signal-to-noise ratio (PSNR) value and structural similarity index (SSIM) value. These two indicators are often used to evaluate the image quality [19]. The larger the PSNR, the smaller the distortion of the output image, and the larger the SSIM, the higher the similarity between the output image and the label image. As shown in Figures 7 and 8, there is a high similarity in the trends of the two curves, first increasing and then stabilizing with a maximum PSNR and SSIM of 26.88 dB and 0.7740, respectively. It indicates that the network gradually converges during the training process and fits the super-resolution task well.

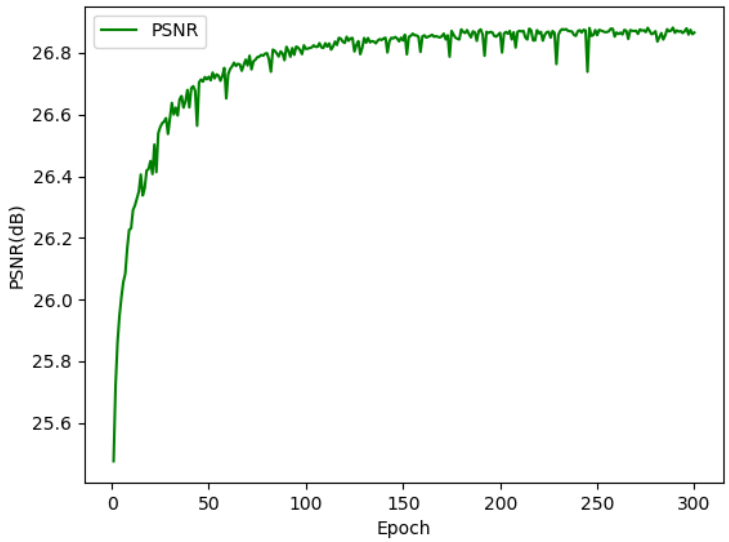

Figure 7. Peak signal-to-noise ratio (PNSR) curve of the validation set.

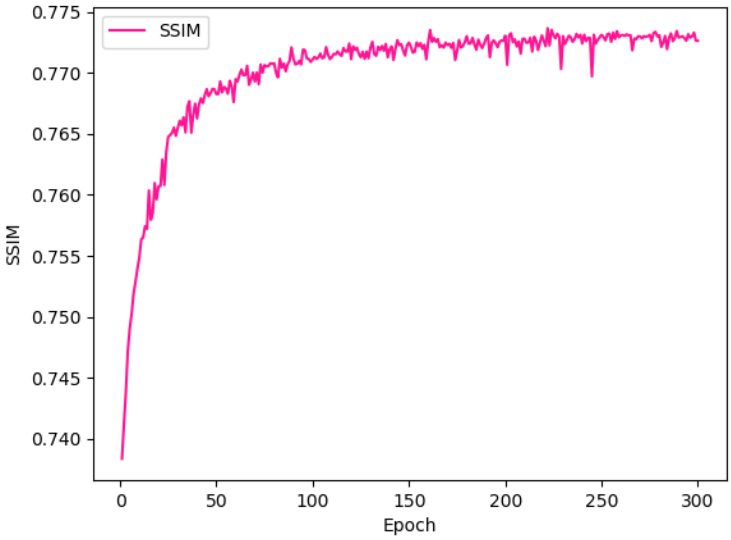

Figure 8. Structural similarity index (SSIM) curve of the validation set.

After training, we tested the trained network with images outside the training and verification sets. We input an image of 200 × 200 pixels into the network to generate an output image of 400 × 400 pixels and compared the output image with a ground truth 400 × 400 pixels image, as shown in Figure 9. By observing the region of interest (depicted by the red box), it can be seen that the output image resolution achieves a significant improvement, particularly at the position indicated by the yellow arrow, where the shape of the cell, which was almost unobservable in the input image, is clearly distinguishable in the output image. In addition, the output image is also comparable to the ground truth image. Processing images using the neural network also produces a large FOV. Assume that the true FOV that can be achieved at 2000× magnification is 27 × 18 μm, whereas the true FOV that can be achieved at 1000× magnification is 54 × 36 μm. The use of a neural network enables the image at 1000× magnification to have the same resolution as the image at 2000× magnification, but with a larger FOV. The output image is then input into the network again, and an output image of 800 × 800 pixels is generated. Compared with ground truth 400 × 400 pixels image, it is clear that the image resolution is further improved. The influence of noise is also reduced to a certain extent. The image therefore appears cleaner.

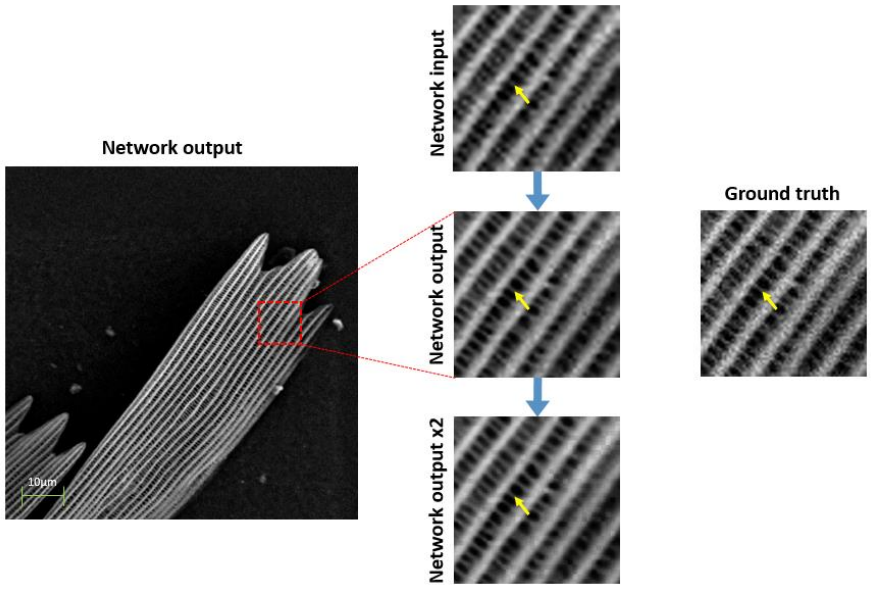

Figure 9. Comparison of neural network output and ground truth.

Next, we compared the trained network with traditional image interpolation methods. We still selected the image data outside the training and verification sets for testing and put 300 × 300 pixel images into the network to generate 600 × 600 pixel output images. At the same time, we used the nearest neighbor interpolation method and cubic interpolation method provided by OpenCV to interpolate and magnify the 300 × 300 pixel images. The three different output images were compared with the ground truth 600 × 600 pixels image. As shown in Figure 10, by calculating the PSNR and SSIM values, we can conclude that the deep learning method has obvious advantages when compared with a traditional image processing method. The values obtained from our method (PSNR = 22.40 dB and SSIM = 0.9451) are higher than those obtained from the NEAREST and CUBIC interpolation methods. (The test phase is numerically calculated for a single image, resulting in a difference in performance from the average value during verification).

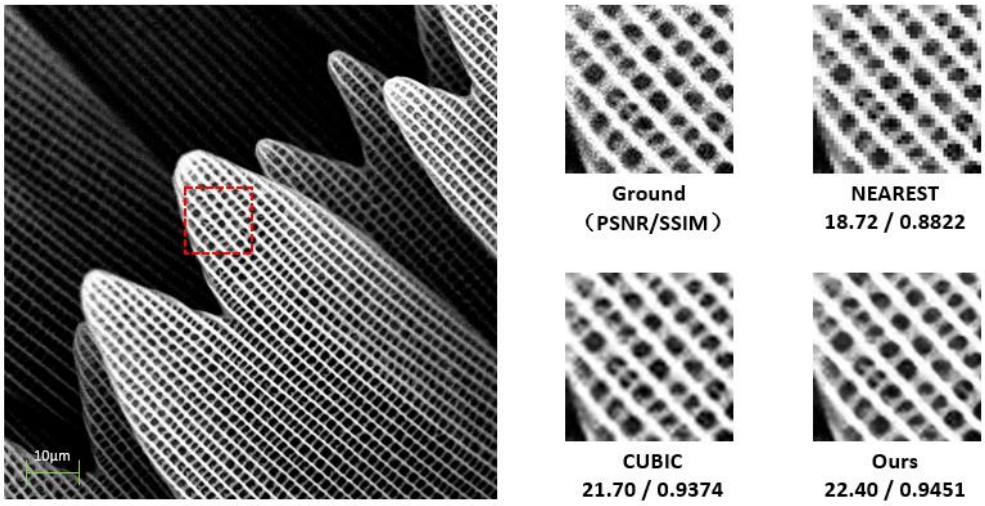

Figure 10. Comparison of output images of deep neural network and traditional interpolation processing methods.

5. Conclusions and Future Work

This study investigates a super-resolution method based on deep learning that can effectively improve the image resolution and the image quality without having to change the physical structure

of the applied FESEM. This deep learning approach can generate an improved image very quickly. It only takes an average of ~0.5 s to output an improved HR image with a large FOV even using a laptop computer.

This deep neural network structure can also be applied to other imaging devices. Training the network with different datasets enables it to complete the super-resolution task of that corresponding device, thus demonstrating excellent expandability.

Future research involves expanding the variety of network structures in our deep neural network such as DenseNet [20], SENet [21], and GAN [22]. These networks exhibit a superior performance in the imaging field, and can optimize end-to-end image generation.

Author Contributions: Z.G. performed the constructive discussions and provided the experimental environment. W.M. contributed to experiments and wrote the manuscript. S.H. and P.H. contributed to manuscript preparation and assistant. C.L. contributed to the conception of the study and provided funding.

Funding: This work was supported by National Natural Science Foundation of China (NSFC 61905021) The Opening Foundation of State Key Laboratory of Millimeter Waves: K202008. Basic Research Freedom Exploration Project of Shenzhen (JCYJ20180305164708625).

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

References

1. Folk, R.L. SEM Imaging of Bacteria and Nannobacteria in Carbonate Sediments and Rocks. *J. Sediment. Res.* **1993**, *63*, 990–999.
2. Perovic, D.D.; Castell, M.R.; Howie, A.; Lavoie, C.; Tiedje, T.; Cole, J.S.W. Field-emission SEM imaging of compositional and doping layer semiconductor superlattices. *Ultramicroscopy* **1995**, *58*, 104–113. [CrossRef]
3. Blazek-Welsh, A.I.; Rhodes, D.G. SEM imaging predicts quality of niosomes from maltodextrin-based proniosomes. *Pharm. Res.* **2001**, *18*, 656–661. [CrossRef] [PubMed]
4. Chen, C.Y.; Klette, R. Image stitching-comparisons and new techniques. *Lect. Notes Comput. Sci.* **1999**, *1689*, 835.
5. Nan, J.; Luo, W. Quantum image scaling using nearest neighbor interpolation. *Quantum Inf. Process.* **2015**, *14*, 1559–1571.
6. Gribbon, K.T.; Bailey, D.G. A Novel Approach to Real-time Bilinear Interpolation. In Proceedings of the IEEE International Workshop on Electronic Design, Perth, Australia, 28–30 January 2004.
7. Gao, S.; Viktor, G. Bilinear and bicubic interpolation methods for division of focal plane polarimeters. *Opt. Express* **2011**, *19*, 26161–26173. [CrossRef] [PubMed]
8. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef] [PubMed]
9. Smirnov, E.A.; Timoshenko, D.M.; Andrianov, S.N. Comparison of Regularization Methods for ImageNet Classification with Deep Convolutional Neural Networks. *Aasri Procedia* **2014**, *6*, 89–94. [CrossRef]
10. Gatys, L.A.; Ecker, A.S.; Bethge, M. Image Style Transfer Using Convolutional Neural Networks. In Proceedings of the Computer Vision & Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
11. Deng, L. Convolutional Neural Networks for Speech Recognition. *IEEE/ACM Trans. AudioSpeechLang. Process.* **2016**, *22*, 1533–1545.
12. Rie, J.; Tong, Z. Effective Use of Word Order for Text Categorization with Convolutional Neural Networks. *arXiv* **2014**, arXiv:1412.1058.
13. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
14. Jin, K.H.; McCann, M.T.; Froustey, E.; Unser, M. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Trans. Image Process.* **2017**, *26*, 4509–4522. [CrossRef] [PubMed]
15. Zhang, H.; Casaseca-de-la-Higuera, P.; Luo, C.; Wang, Q.; Kitchin, M.; Parmley, A.; Monge-Alvarez, J. Systematic infrared image quality improvement using deep learning based techniques. In Proceedings of the Remote Sensing Technologies and Applications in Urban Environments, Edinburgh, UK, 26 October 2016.

16. Wang, S.; Su, Z.; Ying, L.; Peng, X.; Zhu, S.; Liang, F.; Feng, D.; Liang, D. Accelerating magnetic resonance imaging via deep learning. In Proceedings of the IEEE 23th International Symposium on Biomedical Imaging, Prague, Czech Republic, 13–16 April 2016.
17. Rumelhart, D.E. Learning Representations by Back-Propagating Errors. *Cogn. Modeling* **1986**, *5*, 1. [CrossRef]
18. Zhao, H.; Gallo, O.; Frosio, I.; Kautz, J. Loss Functions for Neural Networks for Image Processing. *arXiv* **2015**, arXiv:1511.08861.
19. Horé, A.; Ziou, D. Image quality metrics: PSNR vs. SSIM. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010.
20. Gao, H.; Zhuang, L.; Weinberger, K.Q. Densely Connected Convolutional Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
21. Jie, H.; Li, S.; Gang, S. Squeeze-and-Excitation Networks. In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, Salt Lake City, UT, USA, 18–23 June 2018.
22. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the International Conference on Neural Information Processing Systems, Massachusetts, USA, 10 June 2014; MIT Press: Cambridge, UK.