Noise Reduction of SEM Images using U-net with SSIM Loss Function

Koshiro Nagano¹, Yoshiharu Mukouyama², Takashi Nishimura³, Hiroyuki Fujioka⁴,
Kenji Watanabe⁵, Takio Kurita⁶, and Akinori Hidaka²

¹Graduate School of Science and Engineering, Tokyo Denki University, Japan

²School of Science and Engineering, Tokyo Denki University, Japan

³Osaka Research Institute of Industrial Science and Technology, Japan

⁴Department of System Management, Fukuoka Institute of Technology, Japan

⁵Department of Information Technology and Human Factors, National Institute of Advanced Industrial Science and Technology, Japan

⁶The Graduate School of Advanced Science and Engineering, Hiroshima University, Japan

E-mail: 20rmu07@ms.dendai.ac.jp

Abstract

The image-to-image translation networks, such as U-net [1] or Pix2pix [2], are known to be able to convert input images into different images where the image quality is improved or desired semantic information hidden in the input images are extracted. Several types of research based on such image translation networks have been carried out to realize noise removal systems that convert low-quality images taken with a low-performance microscope into high-quality images taken with a high-performance microscope [3,4,5].

In this paper, we develop denoising and deblurring methods to improve the image quality taken by the conventional scanning electron microscope (SEM) as the level of the image quality taken by the field emission (FE) SEM. In order to realize such methods, we utilize Pix2pix and U-net as the image denoiser.

We compare the results of each image denoiser qualitatively and quantitatively. We show that the images generated by the conventional U-net [6] are apt to be slightly but entirely blurred, and the generative adversarial networks (GAN) [7] comprising a part of Pix2pix has a risk to inappropriately modify image details. Hence, we propose and evaluate U-net using Structural similarity (SSIM) loss function. We show that SSIM U-net can avoid a slight blur caused by the conventional U-net with fewer falsification than Pix2pix.

1 Introduction

The machine learning-based artificial intelligence technologies have been remarkably developed in this decade. Especially, deep neural networks (DNNs) for the tasks of image understanding and image generation have been rapidly developed for a variety of purposes. There is large agreement that the image-to-image translation networks such as U-net [1] or Pix2pix [2] can convert an input image to another image that has improved quality or changed semantic information.

As a typical approach using such image translation networks, there are researches about the super-resolution of low-resolution images and quality improvement for low-

quality images. Among such approaches, some researches have been carried out to realize a system that converts low-quality images taken with a low-performance microscope into high-quality images taken with a high-performance microscope, i.e., noise removal system [3,4,5]. For example, such networks will be able to convert the low-quality image taken by a conventional scanning electron microscope (SEM) into the one that looks like taken by a field emission (FE) SEM.

In this research, we develop several noise removal systems to improve the quality of images taken by SEM. In order to realize such systems, we utilize Pix2pix and U-net as the image denoiser. Our system can improve several kinds of degradation factors of image quality, i.e., low dynamic range, Gaussian blur, and salt-and-pepper noise, from the level of conventional SEM images to the level of FE-SEM images.

We conducted several experiments to reveal the advantages and disadvantages of our methods; For example, it is known that the images generated by conventional U-net are apt to be slightly blurred because of the property of the mean squared error (MSE) loss function. We show that the restored results of SEM images by using conventional U-net could also be slightly but entirely blurred. Meanwhile, the generative adversarial network (GAN) [7] comprising a part of Pix2pix has a risk to inappropriately modify image details which indicate physical, chemical, or biological characteristics caught up by microscopies, by prioritizing to improve image quality over to keep such important details. We also show that the restored results of Pix2pix contain several fabricated particles. Furthermore, as an alternative solution which has fewer drawbacks, we propose U-net using Structural similarity (SSIM) loss function. We show that SSIM U-net could avoid a slight blur caused by the conventional U-net with fewer falsification than Pix2pix.

The remaining part of this paper is constructed as follows: Sec. 2 introduces existing image-to-image translation networks related to this paper. Sec. 3 describes our proposed noise removal system for SEM images based on U-net with SSIM loss function. Sec. 4 shows that the experiment results to evaluate U-net and Pix2pix. Finally, we conclude in Sec. 5.

Figure 1 The network architecture of U-net [1].

2 Related Deep Neural Networks

In this section, we introduce several existing DNNs which are related to this paper.

2.1 U-net

U-net is a kind of the image-to-image convolutional neural networks (CNNs), and the studies of the network received two prizes at ISBI 2015 [1]. The network excels in solving the challenge of segmenting the image pixels on semantic meaning. The network’s name “U-net” originated in that the network architecture looks like the shape of the “U” character. (Fig.1)

The network can be trained end-to-end from using a labeled training dataset. U-net does not have fully connected layers. It consists of a contracting path to capture context and a symmetric expanding path to generate the output by using the captured context. The contracting path follows the typical architecture of CNNs such as AlexNet [8], and it downsamples feature maps to extract semantic information from the overall image while abstracting location information. In contrast, every step in the expanding path upsamples feature maps.

In each level of the expanding path, the feature maps are made by concatenating two types of maps which came from different paths; The one is extracted from the up-convolution stream, and the other is directly copied from the layer at the same level in the contracting path. The path directly came from the contracting path is called the skip connection. While the former maps contain rich semantic information, they only retain coarse location information. Therefore, the latter maps are combined to complement precise location information. U-net is often utilized to analyze the medical images where the accurate location information of objects is important.

2.2 GAN

The generative adversarial network (GAN) belongs to the unsupervised neural network that commits to generating the image without using the labeled datasets. Generally, GAN is constructed from two characteristic internal models; The one is a generative model to grab probability distribution of given data and to generate fake data, and the other is a discriminative model which estimates the probability that a given sample is actual data or not (i.e, fake data generated by the generative model).

In the training procedure, both models are simultaneously optimized; The generative model is trained to maximize the probability that the discriminative model makes misjudges, while the discriminative model is trained to minimize the probability of such misjudges. Such adversarial training allows the generative model to generate elaborate fake data. The competitive situation is often compared to the relationship between the counterfeiter and the crime lab expert. The better the experts spot fake bills, the more the counterfeiter will be able to make realistic counterfeit bills. This repeated diligent application leads to the counterfeit bills that are difficult for the expert to discriminate real or fake.

2.3 Pix2pix

Pix2pix is a kind of the image-to-image translation network based on GAN and supervised training. Given a paired image dataset $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1:N}$, where $\mathbf{x}_i$ is a source image before translation and $\mathbf{y}_i$ is a target image, Pix2pix can construct the image translation map $\hat{\mathbf{y}} = G(\mathbf{x})$ which resembles the input image $\mathbf{x}$ to the target image $\mathbf{y}$.

The network architecture of Pix2pix is illustrated in Fig. 2. It consists of the following two models: A generative model such as U-net, and a discriminative model. Each model performs the same role as GAN’s internal model. The

generator G produces the image $G(x)$ to trick the Discriminator D . Thus the image and the label image: y are not equal but they are very much alike.

Figure 2 The network architecture of Pix2pix [2].

3 Method

The image-to-image translators such as U-net or Pix2pix will be useful in various research field using any type of image data. We can, for example, use them for super-resolution or denoising of microscopic images, if we can prepare appropriate training datasets which are a lot of pairs of an input image and a teacher image of which the desired expression of the input image. However, such a paired dataset of microscopy images is generally expensive to prepare or hard to obtain; For example, we may have to take photos of the same sample by different microscopies in almost completely the same scale, location, and temporal state.

Meanwhile, for the image denoising problem, there is a useful way to obtain paired training datasets at a low cost; It is easy to spoil the high-quality images by adding several kinds of artificial image noises. In this case, we can use noised images as the input I_i and original (high-quality) images as the teacher T_i . Such an approach (called noising-and-denoising) has shown promising results for deblurring [3,6] or denoising [4,5,6]. Therefore, in this research, we also develop denoising and deblurring DNN based on the noising-and-denoising approach.

3.1 Data Making

Similar to [6], we made the training images for DNN as follows. We used the pair of SEM images to train DNN; The one is taken by conventional SEM which is relatively cheap ($\sim 50,000$ USD) and gives low-quality images, and the other is taken by field emission (FE-)SEM, which is relatively expensive ($\sim 500,000$ USD) and gives high-quality images. We refer to the former and the latter images as LQ SEM image and HQ SEM image, respectively. Fig. 3 shows a brief procedure of data making and DNN training [6]. The HQ SEM image is used to obtain a pseudo-LQ SEM image, which is artificially noised and blurred to downgrade its image quality into the level of the LQ SEM image. The intensities of the salt-and-pepper noise and the Gaussian blur which should be added on the HQ image are automatically estimated by the same method as in [6]. Then, we train U-net or pix2pix by using the pseudo-LQ image as the input and the HQ image as the teacher signal. As the HQ image set, we collected 704 image patches, where each patch is 256×256 pixels, from a whole HQ SEM image which is 1440×1008 pixels. The trained DNN can remove Gaussian blur and salt-

and-pepper noise from not only the pseudo-LQ SEM images but also from actual LQ SEM images because it can expect that the intensity of the blur and the noise appeared in the actual LQ image is the same level as the pseudo-LQ image.

Figure 3 The proposed data making procedure.

3.2 U-net using SSIM as Loss Function

As described in Sec. 4, the outputs of Pix2pix are as clear as the HQ-images. However, those results sometimes contain several falsifications, i.e., erasing or creating the small particles. We consider that the shortcoming of Pix2pix is caused by the property of GAN module. The adversarial training between the generator and the discriminator probably spur the generator on the falsification. If there are "doubtful" pixels that are difficult to distinguish whether a real particle or noise in the input image, Pix2pix can freely choose one of the following three ways: #1. clearing the pixels and making them as the background, #2. illustrating them as particles, or #3. keeping them as is. Unfortunately, the ways #1 or #2 are good choices for GAN because these ways can reduce GAN's loss function more than the way #3.

Meanwhile, U-net is known as a strong image-to-image translator without the GAN module. However, as also described in Sec. 4, the denoising and deblurring results of the conventional U-net were still slightly blurred sometimes. Such results may be caused by the property of the mean squared error (MSE) loss function. To overcome these drawbacks, in this paper, we propose U-net with Structural similarity (SSIM) loss function for denoising and deblurring of SEM images.

Given a pair of a source image and a target image, SSIM can be calculated as follows: At first, we crop the rectangle areas $\mathbf{x}$ on the source image and $\mathbf{y}$ on the target image (Fig.4). Next, we evaluate the average, variance, and covariance of the pixel values in the rectangles by sliding the rectangles on the images. The symbols $\mu_{\mathbf{x}}$ and $\mu_{\mathbf{y}}$ denote the mean of pixel values in $\mathbf{x}$ and $\mathbf{y}$, respectively. Also

σ_x and σ_y denote the variance of pixel values in $\mathbf{x}$ and $\mathbf{y}$, respectively. The covariance matrix σ_{xy} is calculated as $\sigma_{xy} = \mu_{xy} - \mu_x\mu_y$ where μ_{xy} denotes the mean of pixel values in $\mathbf{x}$ and $\mathbf{y}$. Then, SSIM is defined as follows:

$$SSIM = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}. \quad (1)$$

Since SSIM is based on the difference between local image patches, it will therefore be more sensitive to image changes within a local area than MSE.

Figure 4 The rectangle of SSIM.

4 Experiments

In this study, we use the images of gold nanoparticles fabricated on glassy carbon taken by conventional SEM or FE-SEM [9]. The images are divided into four types (Data1~4) according to the strength of a sulfuric acid solution and the deposition time of the gold nanoparticles in the solution. In each experimental condition, the size and the shape of fabricated particles are slightly different (Fig.5). We denote images of Data1 taken by a conventional SEM and FE SEM as H_1 and L_1 , respectively. We use the same symbols for Data2 to Data4 (H_2 , L_2 , and so on). The concentrations of sulfuric acid solution in Data1 to Data3 are constantly 1.97×10^4 ppb, and Data4 is 1.97×10^3 ppb. The deposition time is 1 hour for Data1, 2 hours for Data2, and 4 hours for Data3 and Data4.

Whole image sizes of the conventional SEM and FE-SEM are 2560×1730 and 1440×1008 pixels, respectively. We cropped 256×256 pixel patches by the stride of 80 pixels from the whole images. These 256×256 pixel patches are used for the training and evaluation of DNNs. In this section, we refer to U-net which is trained by using MSE loss and SSIM loss as "MSE U-net" and "SSIM U-net", respectively.

4.1 Comparison between MSE U-net and SSIM U-net

In this section, we describe the restored results of the pseudo-LQ SEM images, that are artificially blurred and noised images. In Fig. 6, we can see that the restored result of MSE U-net is still slightly blurred. On the other hand, the result of SSIM U-net is clearer and sharper than the result of MSE U-net.

In order to quantitatively evaluate the intensity of blurring, we calculate the blur metric [10,11] from the restored images of MSE U-net and SSIM U-net. As can be seen in Table 1, the blur metric of SSIM U-net for Data1 is lower than

MSE U-net. For Data2 and Data3, the intensities of blurring seem to be the same level. This is probably due to the fact that the pseudo-LQ images for Data2 and Data3 are not as strongly blurred as Data1.

For Data4, SSIM U-net showed the worse result. However, the blur intensity of Data4 (< 0.6) is relatively smaller than the other data. As can be seen in Data4 of Fig. 7, since the gold particles were not fabricated well, each particle is quite small. In this case, the effect of salt-pepper-noise becomes relatively larger. It may have made the noise removal including blurring difficult.

Table 1 The blur metric of the restored (pseudo-HQ) image. The higher value indicates a stronger blur.

	MSE U-net	SSIM U-net
Data1	0.7709	0.7487
Data2	0.6976	0.7013
Data3	0.6603	0.6625
Data4	0.4696	0.5840

4.2 Comparison between Pix2pix Image and SSIM U-net Image

In this section, we describe the comparison results of SSIM U-net and Pix2pix for the pseudo-LQ SEM images. As can be seen in the results of Data 1 in Fig. 7, the output of Pix2pix includes several clear fake particles (indicated by yellow circles) that do not exist in the ground truth image. In contrast, there are few such fake particles in the output of SSIM U-net (Fig. 7).

4.3 Restored Results of LQ Images

Finally, we describe restoration results of the images taken by the conventional SEM (that are the *actual* LQ images). The trained DNNs can remove the Gaussian blur and the salt-and-pepper noise from not only the pseudo-LQ SEM images but also from the *actual* LQ SEM images because the intensities of the blur and the noise appeared in the pseudo-LQ SEM images is set to the same level as the actual LQ images.

Fig.8 shows the denoised and deblurred results of the LQ SEM images. The pseudo-HQ images which are denoised by SSIM U-net and Pix2pix are shown. Also, for reference, the HQ SEM images are shown to indicate the desired image quality. As can be seen in the results of Data 1 in Fig. 8, the image quality of the restored results seems to be similar to the desired image quality. Considering the original quality of the LQ image is quite low, the result seems to be substantially good. Meanwhile, in Data2 to Data3, there are some blurred artifacts on the edges of the particles.

Note that the results for Data4 seem to be not good. It may be because of a fault in the process of artificial noising and blurring; We can see that the *actual* LQ image (Fig. 5 right) and the pseudo-LQ image (Fig. 6 left end) of Data4 have slightly different looking. If the pseudo-LQ images do not correctly correspond to the *actual* LQ images, our networks will not be able to function correctly.

5 Conclusions

In this study, we showed that our trained DNNs (U-net or Pix2pix) can remove the Gaussian blur and the salt-and-pepper noise from given LQ SEM images. By using simulated images, we revealed that the images generated by MSE U-net could be slightly but entirely blurred if the LQ images contain a strong blur. Also, we showed that the image generation by Pix2pix easily causes the fabrication of particles in SEM images.

As an alternative solution, we proposed SSIM U-net. We showed that SSIM U-net can avoid a slight blur caused by MSE U-net with fewer falsification than Pix2pix. However, if the LQ images do not have a strong blur, SSIM U-net might not exceed MSE U-net.

In the denoised and deblurred results of the real LQ images, the networks substantially improved the image quality of Data1. For Data2 and Data3, while the networks seem to reduce the blur and the noise, there are several blurred artifacts on the edges of these particles.

Meanwhile, our networks could not function well for Data4. The results may be caused by the slight difference between the real LQ SEM images and the pseudo-LQ SEM images of Data4. In such a situation, the real LQ images of Data4 will be seemed as "unknown" or "unlearned" sample for the networks trained by using the pair of pseudo-LQ image and HQ image of Data4. To handle such a situation, we need to construct a network which has better generalization performance for such unlearned samples. The domain adaptation approach [12] may be helpful to solve this problem.

Acknowledgments

This work was supported by MEXT/JSPS KAKENHI Grant Number JP18H04745 "Resonance Bio" and Research Institute for Science and Technology of Tokyo Denki University Grant Number Q20J-02 / Japan.

References

- [1] O. Ronneberger, et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation", MIC-CAI, Springer, LNCS, Vol.9351, pp. 234-241, 2015.
- [2] P. Isola et al., "Image-to-Image Translation with Conditional Adversarial Networks", CVPR2017, pp. 1125-1134, 2017.
- [3] M. Hradiš et al., "Convolutional neural networks for direct text deblurring", BMVC2015, pp. 1-13, 2015.
- [4] H. Fujioka et al., "Understanding Deformation Motion of Colloidal Nanosheets from CLSM Images Using Deep Learning-Based Approach", ICARCV2018, pp.192-197, 2018.
- [5] H. Fujioka et al., "Detecting Nanosheet Objects from Noisy CLSM Images Using Deep Learning Approach", Key Engineering Materials, Vol. 804, pp. 11-15, 2019.
- [6] A. Hidaka et al., "Image Denoiser for Microscopic Images Based on Noising and Denoising Approach", RBIS2019, pp.72, 2019.
- [7] I.J. Goodfellow et al., "Generative Adversarial Nets", NIPS2014, pp. 2672-2680, 2014.
- [8] A. Krizhevsky, et al., "ImageNet classification with deep convolutional neural networks", Communications of the ACM, Vol.60 (6), pp. 84-90, Retrieved 2017.
- [9] Y. Mukouyama et al., "Fabrication of Uniformly Sized Gold Nanoparticles on Glassy Carbon by Simple Electrochemical Method", Journal of The Electrochemical Society, Vol.166 (13), pp. 669-675, 2019.
- [10] F. Crete et al., "The blur effect: Perception and estimation with a new no-reference perceptual blur metric", SPIE, Vol.6492, pp. 64920, 2007.
- [11] Do Quoc Bao, "Image Blur Metric", MATLAB Central File Exchange, <https://www.mathworks.com/matlabcentral/fileexchange/24676-image-blur-metric>, Retrieved 2020.
- [12] E. Tzeng, et al., "Adversarial Discriminative Domain Adaptation", CVPR2017, pp. 7167-7176, 2017.

Figure 5 Example images of gold nanoparticles fabricated on glassy carbon in different experimental conditions.

Figure 6 Comparison of the restored results of the pseudo-LQ SEM images. The result of MSE U-net is slightly but entirely blurred.

	Input (Pseudo-LQ Image)	Ground Truth (HQ Image)	MSE U-net	SSIM U-net (Ours)	Pix2pix
Data 1					
Data 2					
Data 3					
Data 4					

Figure 7 Restored results of the artificially blurred and noised SEM images.

Figure 8 The denoised and deblurred results of the actual LQ Images.