INVESTIGATION OF DIFFERENT IMAGE SUPER RESOLUTION METHODS ON PAIRED ELECTRON MICROSCOPIC IMAGES

A Thesis

by

JIAXI XU

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Yu Ding
Committee Members,	Satish Bukkapatnam
	Xiaoning Qian
Head of Department,	Lewis Ntaimo

December 2020

Major Subject: Industrial & Systems Engineering

ABSTRACT

This thesis is concerned with investigating super-resolution algorithms and solutions for handling electron microscopic images. Please note two main aspects differentiating the problem discussed here from those considered in the literature. The first difference is that in the electron imaging setting, a pair of physical high-resolution and low-resolution images is used, rather than a physical image with its downsampled counterpart. The high-resolution image covers about 25% of the view field of the low-resolution image, and the objective is to enhance the area of the low-resolution image where there is no high-resolution counterpart. The second difference is that the physics behind electron imaging is different from that of optical (visible light) photos. The implication is that super-resolution models trained by optical photos are not effective when applied to electron images. Focusing on the unique properties, a global and local registration method is devised to match the high- and low-resolution image patches and different training strategies are discussed for applying deep learning and non deep learning based super-resolution methods to the paired electron images.

This thesis investigates the uniqueness of the super-resolution problem on paired electron microscopic images. After extensive experimentation and comparison on 22 pairs of electron images, it is now believed that the self-training strategy, in which the training images come from the same image pair of the test set, leads to better super-resolution outcomes, despite the relatively small training data size. Deep learning-based super-resolution methods show the best performances, whereas a revised paired library-based non-local mean method shows advantage in training time and interpretability.

Paired images super-resolution has important implications in many research areas. Paired electron images are rather common in scientific experiments, especially in material and medical research. Due to the destructive imaging process while using electron sources, researchers tend to use low-energy beams or subject the samples to a short duration of exposure to protect the sample. As a consequence, low-resolution images are generated. Super-resolution methods, which can subsequently boost these low-resolution images to a higher resolution, are much desired in scientific researches using electron imaging.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professor Yu Ding [advisor] and Professor Satish Bukkapatnam of the Department of Industrial & Systems Engineering and Professor Xiaoning Qian of the Department of Electrical & Computer Engineering.

The image data analyzed in the thesis was provided by Professor Yu Ding. The analyses depicted in Chapter 4 were conducted in part by Yanjun Qian of the Department of Statistical Sciences and Operations Research, Virginia Commonwealth University and were submitted in 2020 in an article listed in the References.

All other work conducted for the thesis was completed by the student independently.

Funding Sources

Graduate study was supported by AFOSR DDDAS program grants FA9550-18-1-0144 and Texas A&M X-grant program.

NOMENCLATURE

OGAPS	Office of Graduate and Professional Studies at Texas A&M University
TAMU	Texas A&M University
EM	Electron Microscopic
HR	High-Resolution
LR	Low-Resolution
SR	Super-Resolution
SEM	Scanning Electron Microscope
TEM	Transmission Electron Microscope
NLM	Non-Local Mean
LB-NLM	Library-Based Non-Local Mean
ScSR	Sparse-coding Super-Resolution
SRSW	Super-Resolution via Sparse Weighting
VDSR	Very-Deep Super-Resolution
EDSR	Enhanced Deep-residual networks Super-Resolution
RCAN	Residual Channel Attention Networks
PSNR	Peak Signal-to-Noise Ratio
SSIM	Structural Similarity
MSE	Mean Squared Error
HPRC	High Performance Research Computing

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION[*]

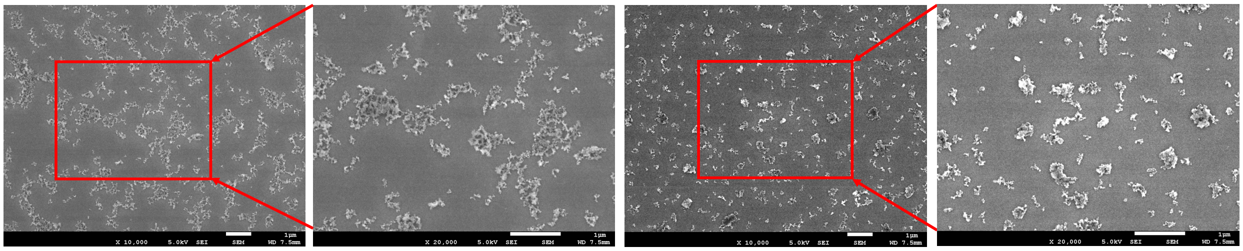
In this thesis, we consider an image processing problem encountered in nanomaterial characterization. Material science researchers capture two-resolution electron microscopic (EM) images independently from the same specimen: a high-resolution (HR) image of $M \times N$ pixels, denoted by $\mathbf{I}_h$, and a low-resolution (LR) image, denoted by $\mathbf{I}_l$. The LR image has the same amount of pixels of the HR image but half of its resolution.

Figure 1.1 demonstrates two pairs of such EM images, both obtained by a scanning electron microscope (SEM). Each pair of images is obtained by the same SEM in one experimental setting but through two actions. First, the SEM is set at a low magnification level and takes the low-resolution image. Then, with the same sample still in the specimen platform, the SEM is adjusted to a higher magnification level, i.e., it is zoomed in, and takes the high-resolution image. The view fields of the two images overlap completely, or more precisely, the high-resolution image covers a smaller subset of the view field of the low-resolution image. The overlapping areas in the LR images in Figure 1.1 are marked by the red rectangles. The objective is to develop a super-resolution (SR) method for reconstructing an HR image of $2M \times 2N$ pixels over the whole area that is covered by the LR image. The essence of the task is to enhance the area of the low-resolution image where there is no high-resolution counterpart. If a method can accomplish this research goal, material scientists can effectively survey a bigger area with imaging quality comparable to HR images but with less dense sampling.

HR images are desired for the purpose of material characterization because they capture and reveal fine structures of nanomaterials [1, 2, 3, 4]. But it is time consuming to capture HR images. While using a SEM or a transmission electron microscope (TEM), the images are created by an electron beam rastering through the material, so that the time cost will be at least proportional to the number of pixels. An equally important consideration is that the electron beam of an EM may

---

[*] © 2020 IEEE. Reprinted, with permission, from Y. Qian, J. Xu, L. F. Drummy and Y. Ding, "Effective Super-Resolution Methods for Paired Electron Microscopic Images," *IEEE Transactions on Image Processing*, vol. 29, pp. 7317-7330, 2020.



LR SEM Image HR SEM Image LR SEM Image HR SEM Image

Figure 1.1: Two pairs of LR and HR SEM images. The red rectangles in the LR images are the areas corresponding to the HR images.

damage the subtle structure of certain materials. Material scientists want to avoid dense sampling in electron imaging if at all possible. An effective SR approach, if available, can be of a great help to high-resolution electron imaging of materials.

In recent years, numerous SR methods have been proposed and reported [5, 6, 7, 8]. We note two important differences, distinguishing the problem at hand from those considered in the literature. The first difference is that we have a pair of LR and HR images, both physical and obtained independently, rather than a physical HR image with its downsampled version. It is our understanding that most of the existing SR approaches in their default settings take the downsampled version of the HR images as the LR image inputs in their training. The second difference is that the physics behind electron imaging is different from that of optical photos taken under visible light. The implication is that super-resolution methods trained by optical photos are not going to be effective when applied to electron images. We will provide quantitative evidence in Chapter 2 to support our claim.

Focusing on these unique properties of our problem, we first examine how the existing methods perform while training with the physically captured image pairs. We test the two sets of popular SR methods: the sparse-coding based SR [9, 10] and deep-learning based SR [11, 12, 13]. While the sparse-coding methods fail to yield satisfying results, we find that the deep-learning based approaches demonstrate a good degree of adaptability to our problem. Qian et al. [14] propose a simpler SR method based on non-local means (NLM) [15, 16]. A bit surprisingly, the NLM method performs rather competitively—as the closest runner-up and only slightly less effective than the deep learning-based SR. The NLM method, on the other hand, is fast to train and has good interpretability.

We investigate different training strategies. We find that the self-training, in which the model is trained by the data from a specific pair of images, attains the most competitive performance for all methods, in spite of the limited number of training samples under such setting. This revelation appears to differ from the conventional wisdom in SR research, which prefers a large size of training samples even though some or many of the training samples are not directly related to the test image. Under self-training, we observe that simpler networks among the deep-learning approaches can produce SR results as competitive as complex networks but the training of the simple networks is much faster. All these observations indicate that the strategy for super resolving the paired EM images is different from those for general SR problems.

The remaining parts of this paper are organized as follows. Chapter 2 reviews the relevant literature and presents our preliminary analysis, which is to demonstrate that a model trained by an external, synthetic dataset is not effective for the paired image problem. In Chapter 3 we explain how to tailor the existing SR methods for the paired EM image problem. Chapter 4 compares the performance of multiple methods and shows the benefits of the self-training strategy. In Chapter 5, we summarize our contributions and discuss possible extensions.

2. LITERATURE REVIEW AND PRELIMINARY ANALYSIS*

While the early SR literature focuses on restoring an HR image with multiple LR images (e.g., those in a short clip of video) [5, 17, 18], the mainstream SR research nowadays is the single-image SR, starting with the seminal work by Freeman et al. [19] nearly twenty years ago. The idea of single-image SR is as follows. HR/LR patches are extracted from a set of training images, and a machine learning model is then built to map the images at the two resolutions. A test LR image will be segmented into overlapping patches, and the corresponding HR patches are to be inferred by the trained model. The HR image over the whole field of view is then reconstructed from these HR patches.

Numerous single-image SR methods for optical images have been proposed using different machine learning models. The neighborhood embedding (NE) algorithms [20, 21, 22] are based on the assumption that the HR and LR patches share similar manifold structures. An HR patch is estimated from the nearest neighbors of its LR counterpart in the manifold. The joint model methods [23, 24] learn a joint HR-LR patch distribution and predict HR images by maximizing the likelihood. The regression-based algorithms [25, 26, 27, 28] fit a regression model to map LR and HR patches and predict the HR patch using the LR patches as the regressors. The sparse-coding super-resolution (ScSR) methods [9, 10, 29, 30] look for a parsimonious dictionary to encode the patches and reconstruct the HR patch from the coefficients of its LR counterpart. In recent years, the deep learning methods have been adopted for achieving single-image SR [11, 12, 13, 31, 32] and image restoration [33, 34, 35]. The deep-learning methods, e.g., very-deep super-resolution (VDSR) [11], enhanced deep-residual networks super-resolution (EDSR) [12] and residual channel attention networks (RCAN) [13], achieve the best performance in recent single-image SR challenges [8, 36].

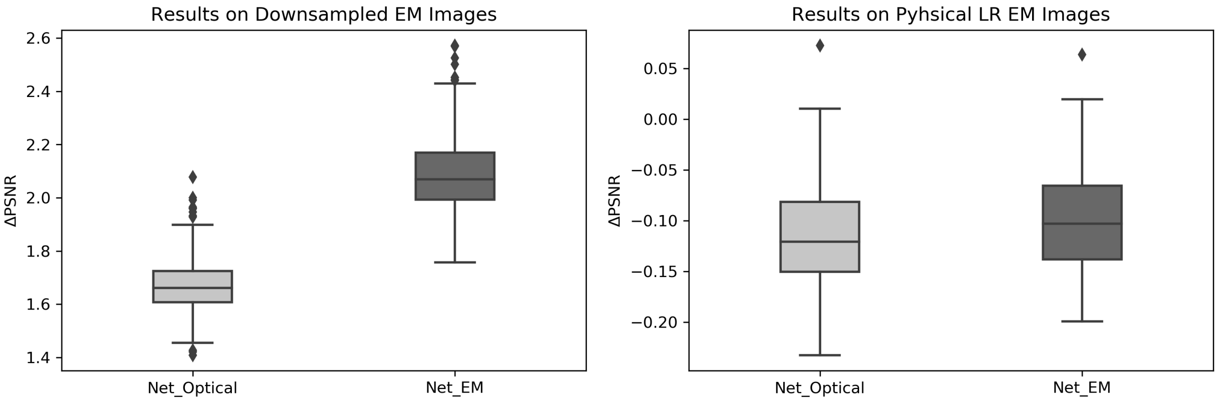
In their default setting, ScSR and deep learning methods usually train their models from high-

---

[*] © 2020 IEEE. Reprinted, with permission, from Y. Qian, J. Xu, L. F. Drummy and Y. Ding, "Effective Super-Resolution Methods for Paired Electron Microscopic Images," *IEEE Transactions on Image Processing*, vol. 29, pp. 7317-7330, 2020.

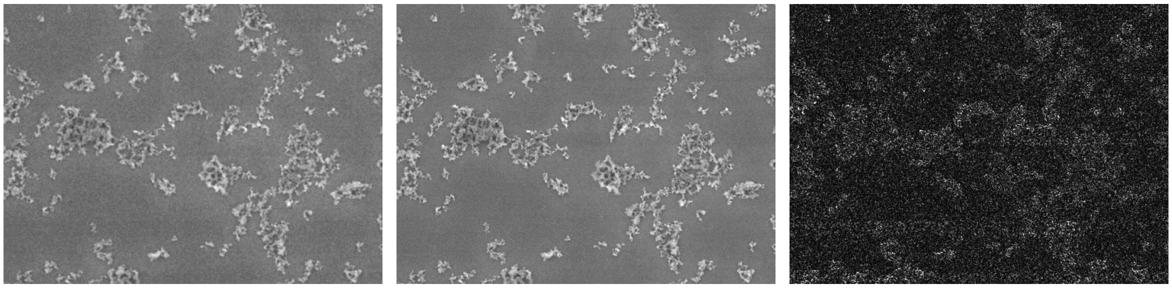
resolution optical images only. An LR image is used but it is synthesized by blurring and downsampling the HR image. We refer to this type of LR images as the synthetic LR images. We design a preliminary experiment to demonstrate that this default setting is ineffective when applied to the paired EM images.

We train two networks with 41 layers by VDSR using its default setting in [11]: *Net_Optical* from 539 HR optical images from the IAPR TC-12 Benchmark [37] , and *Net_EM* from 539 HR EM images collected by ourself. Then we test these networks using two datasets: synthetic electron images downsampled from the HR images and the physical LR electron images corresponding to the same HR images. After that, we compare the reconstructed images, presumably enhanced, with the actual HR images and calculate the peak signal-to-noise ratio (PSNR)—a high PSNR indicates a good reconstruction. Our baseline method is the bicubic interpolation [38], which is the most popular algorithm for upsampling an LR image to the pixel amount of the HR images. ΔPSNR is computed as the difference between the PSNR of the image processed by a target method and the PSNR of the same image processed by bicubic interpolation. ΔPSNRs are shown in Figure 2.1 when the two networks are applied to these two datasets.



(a) (b)

Figure 2.1: The performance of VDSR when its two versions, *Net_Optical* and *Net_EM*, are applied to the downsampled and physical LR EM images, respectively.


Physical LR Image

Synthetic LR Image

Absolute Values of the Difference

Figure 2.2: Comparison of a physical LR EM image and the synthetic downsampled image from the common corresponding HR image.

One can make two observations from Figure 2.1. First, for both datasets, *Net_EM* is more effective than *Net_Optical*, highlighting the difference between optical and EM images. Second, although the two networks both work well on downsampled EM images (left panel), they perform worse relative to bicubic interpolation (right panel) when applied to the physical LR images. The message is that material scientists cannot simply grab an existing pre-trained SR model for processing the paired EM images. When we tried the sparse-coding SR methods [9, 10] or other deep-learning SR [12, 13] with their default setting, which use synthetic LR images, the resulting SR models are similarly ineffective.

In Figure 2.2, we compare a physical LR image and a synthetic image, blurred and downsampled from their commonly paired HR image, and highlight their discrepancy. As we see in the right-most plot of Figure 2.2, the difference between the two images is rather pronounced. We believe that the reason of discrepancy is in fact complicated, caused by the noise existing in the HR image, the different contrast levels between the paired images, and/or different natures and degrees of local distortion from individual image-capturing processes.

Although not studied in the SR literature as thoroughly as the single-image SR problems have been, there are in fact some initial attempts on the SR problem involving physically captured LR images. Xu et al. [39] develop a SR approach for blurred LR face and text images with a generative adversarial network (GAN). Then, Xu et al. [40] propose a framework to generate realistic training data from raw images captured by a camera sensor, and improve the SR results from the real LR optical images. Zhang et al. [41] discuss super resolution trained with physical LR images. They improve the traditional deep learning based SR by using raw images from a camera and introducing a new loss function for handling the local distortions. Those works confirm and highlight the drawbacks of training from synthetic images when processing physical images, inspiring us to extend this line of research to paired EM images.

Trinh et al. [10] propose a SR method for paired medical images. To handle the noise in LR images, they store original image pairs to build a library and then solve a sparse representation for an input LR patch to reconstruct its HR counterpart. While accounting for noise in LR images, Trinh et al. [10] still ignore other discrepancies between the image pairs, e.g., the local distortion and differing contrasts. Their reconstruction process is also slow as solving the L_1 optimization for sparse representation is time-consuming.

Sreehari et al. [16] propose one of the first SR methods for EM images. In their approach, a library is built by an HR scan over a small field-of-view of a certain sample. When the LR EM image over a large field-of-view comes, a library-based non-local-mean method (LB-NLM) [15] is applied to the upsampled LR image. After that, the HR image is recovered in a plug-and-play framework by invoking an alternating direction method of multipliers (ADMM) solver [42]. Compared with the SR methods for optical images, Sreehari et al. [16] build the library directly using electron image samples of nanomaterials, rather than unrelated optical images, and consider the noise in HR images explicitly in the plug-and-play framework. However, their algorithm does not include the physical LR images in the library, falling short of mapping the LR and HR patches directly.

Qian et al. [14] propose a simple but effective SR method for the paired EM images, based on the LB-NLM filtering [15, 16]. Qian et al. [14] build a paired library to connect the HR and LR patches from the training images. To include the informative patches for better training results, a clustering method is designed for selecting the representative patches. The last step is a revised library-based non-local-mean (LB-NLM) method that reconstructs the HR images over the whole field of view, using the paired library of representative patches. This last step is one of the key differences between the original LB-NLM method in [16] and that in [14]. As mentioned above, Sreehari et al. [16] do not include the physical LR images in their library, but Qian et al. [14] use the physical LR images explicitly.

3. SUPER-RESOLUTION METHODS FOR PAIRED EM IMAGES*

In this chapter, we proceed with different approaches for handling paired EM images. We apply the current SR methods, specifically the sparse-coding methods [9, 10], deep-learning methods [11, 12, 13] and paired LB-NLM based [14], using the physical EM image pairs as input. To handle the uniqueness of paired EM images, we explore different training strategies. The common preprocessing step for all methods is to register the HR and LR physical images; for that, we devise a global and local registration procedure. The global registration is applied to the whole image, so that this step is common to all SR methods. The local registration is applied to the image patches and thus common only to the sparse coding methods and the paired LB-NLM method. The deep learning methods take the whole images as input to their networks and conduct an end-to-end super-resolution; for them, only is the global registration applied. In Chapter 3.3, we discuss the performance criteria used to evaluate the efficacy of the SR methods. Along with the commonly used PSNR and structural similarity (SSIM) [43], we also introduce some new metrics that we believe can articulate more pointedly the improvement made by the SR methods in the context of material characterization.

3.1 Global and Local Registration

With a pair of HR/LR EM images, $\mathbf{I}_h$ and $\mathbf{I}_l$, as inputs, we upsample $\mathbf{I}_l$ by a factor of two using bicubic interpolation; this produces $\mathbf{I}_u$, a $2M \times 2N$ image. Then a shift transform (x, y) and a rotation transform (θ) are applied to $\mathbf{I}_u$ and the mean squared error (MSE) between $\mathbf{I}_h$ and $\mathbf{I}_u$ are calculated in their overlapping area. We use a grid search to identify (x, y, θ) to globally minimize the MSE. To accelerate the searching process, we first downsample the two images by the same factor and roughly estimate (x, y, θ). Then we refine the estimation by searching its neighborhood using the original images. The registered upsampled image, denoted by $\mathbf{I}_r$, is transformed from $\mathbf{I}_u$

---

*© 2020 IEEE. Reprinted, with permission, from Y. Qian, J. Xu, L. F. Drummy and Y. Ding, "Effective Super-Resolution Methods for Paired Electron Microscopic Images," *IEEE Transactions on Image Processing*, vol. 29, pp. 7317-7330, 2020.

using the optimal global registration parameters.

To handle the local distortions between images, we segment the matched $\mathbf{I}_h$ and $\mathbf{I}_r$ into overlapping patches of size $n \times n$. $\mathbf{P}_h(i, j)$ and $\mathbf{P}_r(i, j)$ denote, respectively, the patches centered at (i, j) in $\mathbf{I}_h$ and $\mathbf{I}_r$. Then we search the neighborhood of (i, j) to find (i^*, j^*) via solving the following optimization problem:

$$\min_{i^*, j^*} \frac{\mathbf{P}_h(i^*, j^*) \cdot \mathbf{P}_r(i, j)}{\|\mathbf{P}_h(i^*, j^*)\|_F \|\mathbf{P}_r(i, j)\|_F}, \tag{3.1}$$

where $\cdot$ denotes the inner product and $\| \ \|_F$ is the Frobenius norm or the entrywise matrix 2-norm. We prefer the use of an inner product over the use of a Euclidean distance to match the two patches as the former is insensitive to the contrast difference between the two images. This criterion become less effective when the patches contain poor texture. Fortunately, the patches containing poor texture are the background patches, which are less important to the mission of super-resolution. We only apply the local registration to the patches with rich texture, which can be selected by deeming the variance of $\mathbf{P}_r(i, j)$ of a patch larger than a certain threshold. For our EM images, we set the threshold as 100. Figure 3.1 presents one example after local registration, where the red arrows illustrate the displacements $(i^* - i, j^* - j)$ between the matched patches in $\mathbf{I}_h$ and $\mathbf{I}_r$. The magnitudes and directions of the displacements vary significantly across the image, showing a complex and irregular pattern of local distortions, which would not have been adjusted by a global registration alone.

3.2 Existing SR Methods Applied to Paired Images

After image registration, we can apply the popular SR methods to the paired EM images. Here we test three main approaches: the sparse-coding methods, the deep learning methods and the paired LB-NLM method.

When using the sparse-coding methods, we decide to remove the back-projection step after the SR reconstruction. The back-projection step was included in the original sparse-coding method under the assumption that by downsampling the SR result, one can get the same image as the LR input. This assumption is not valid for the paired EM images; we articulated this point in Figure

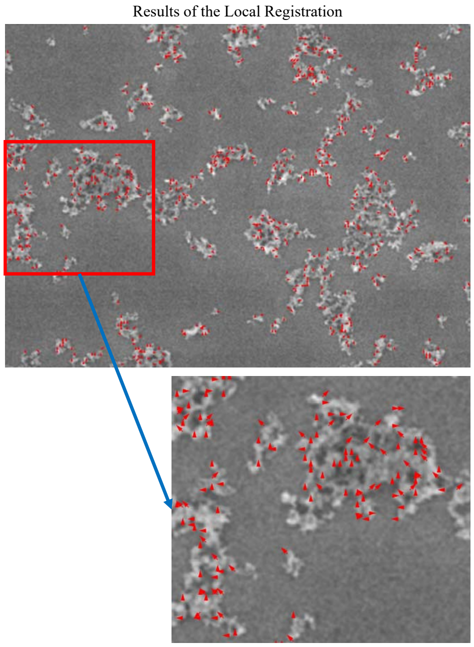


Figure 3.1: The results of the local registration. The bottom figure is magnified from the red rectangle in the top figure, in which the red arrows indicate the displacements $(i^* - i, j^* - j)$ between the matched patches.

2.2. Our test shows that including the back-projection step deteriorates the SR result instead of improving it when the ScSR method is applied to the paired EM images.

When using the deep-learning methods, we are mindful of the small sample size of the paired EM images. The small number of EM images is a result of the expensiveness to prepare material samples and operate electron microscopes. Acquiring *paired* EM images would be even more time-consuming because doing so needs special care and specific experimental setup. In reality, one can expect a handful, to a few dozens at best, of such paired EM images. To prevent overfitting, we adopt two techniques: data-augmentation and early-stopping. A larger dataset is created by flipping each image pair row-wise and column-wise, rotating them by 90, 180 and 210 degrees, and downsizing them by the factors of 0.7 and 0.5. By calculating the accuracy using validation data, we also discover that training achieves the best performance before its 30th epochs and should be stopped accordingly.

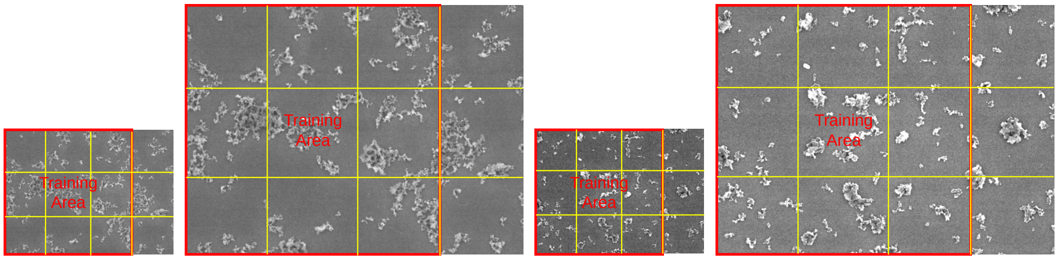
Since the paired LB-NLM method is devised specially for the paired EM images, based on the LB-NLM filtering [15, 16], we implement it without modification.

There is the question of how to train a SR model. The use of external image datasets for training, as done in the current SISR, is not the best practice in handling paired image problems, as shown in our preliminary analysis. Being “external”, it means that the image pairs in the training set are unrelated to the image to be super resolved. That setting is understandable when one only has an LR image without its HR counterpart. For the paired image cases, given the complete overlap, albeit a subset of the view field, between an LR image and its HR counterpart, one would think that a relationship learned directly from this specific pair is the best for boosting the resolution for the rest of the LR image area uncovered by the HR image.

Suppose that we have a total of m_{pr} pairs of SEM images, each of which has an LR image and its corresponding HR image. In this particular study, $m_{\text{pr}} = 22$. The size of both types of images is $1,280 \times 944$ pixels. Through image registration, we identify the overlapping areas of each pair and carve out the corresponding LR image, which is of 640×472 pixels. The $1,280 \times 944$-pixel HR image and the 640×472-pixel LR image are what we used to train the model and do the testing. The non-overlapping area of the LR image is not used in the experimental analysis because there is no ground truth for that area to be tested.

To mimic the practical applications where the SR method is to be applied to the area where there is no corresponding HR images, we partition the LR and HR images in each pair into 3×4 subimages. We treat $m_{\text{pp}}^{\text{tr}}$ subimages as the training images and keep the remaining $m_{\text{pp}}^{\text{ts}}$ subimages unused in the training stage and treat them as the out-of-sample test images. In this study, the number of training images per pair is $m_{\text{pp}}^{\text{tr}} = 9$ and the number of test images per pair is $m_{\text{pp}}^{\text{ts}} = 3$. The size of an HR subimage is 320×314, where the size of an LR subimage is 160×157, still maintaining the 2:1 resolution ratio. The training and test subimages of two SEM image pairs are shown in Figure 3.2.

There are naturally two training strategies. To reconstruct the test subimages from Pair $i = 1, \ldots, m_{\text{pr}}$, we can use the training subimages coming from the same pair to train the model. As



Overlapping Area of Pair 1

Overlapping Area of Pair 2

Figure 3.2: The overlapping areas of two pairs of SEM images. The left 75% is the training area and the right 25% is the test area. The yellow lines partition each image into 3×4 subimages.

such, there will be m_{pr} individual models trained. In the phase of testing, each model is used individually for the specific image pair from which the model is trained. Each model is trained by $m_{\text{pp}}^{\text{tr}}$ pairs of subimages and evaluated on $m_{\text{pp}}^{\text{ts}}$ pairs of subimages. We refer to this strategy as self-training.

Alternatively, we can pool all the training sample pairs together and train a single model. In the phase of testing, this single model is used for reconstructing the test images for all image pairs. We refer to this strategy as pooled-training. Under this setting, there are a total of $m_{\text{pr}} \times m_{\text{pp}}^{\text{tr}}$ pairs of training images and $m_{\text{pr}} \times m_{\text{pp}}^{\text{ts}}$ pairs of test images. In the above example, the training sample size in the pooled-training is 198 pairs of subimages and the test sample size is 66 pairs of subimages, much greater than the sample sizes used in self training.

The conventional wisdom, especially when deep learning approaches are used, is that the $m_{\text{pp}}^{\text{tr}}$ training images, which are nine in this example, are too few to be effective. The popular strategy is to use the pooled training. For the paired EM images, however, we find that using self-training in fact produces the best SR results, despite the relatively small sample size used. We believe this is something unique for the paired EM image problem—the pairing in the images makes using training samples internal to a specific image pair a better option than using more numerous external images. We will present numerical evidences in Chapter 4.

3.3 Performance Criterions for Nanoimages

To measure the performance of an SR method, the standard approach is to consider the HR image as the ground truth, and compare it with the reconstructed image. Two most popular quantitative metrics are the peak signal-to-noise ratio (PSNR) and the structural similarity (SSIM) index. The closer the two images are, the higher PSNR and SSIM will be.

To define PSNR between two images, let us define the MSE first. For two images $\mathcal{A}$ and $\mathcal{B}$ of the same size $M \times N$, the MSE between them is defined as

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (\mathcal{A}(i,j) - \mathcal{B}(i,j))^2. \tag{3.2}$$

Then, PSNR is defined as

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}^2}{\text{MSE}} \right) = 20 \cdot \log_{10} \text{MAX} - 10 \cdot \log_{10} \text{MSE}, \tag{3.3}$$

where MAX is the maximum pixel value of the baseline image, $\mathcal{A}$. When samples are represented with B bits per sample, MAX $= 2^B - 1$.

SSIM is defined as follows,

$$\text{SSIM} = \frac{(2\mu_{\mathcal{A}}\mu_{\mathcal{B}} + c_1)(2\sigma_{\mathcal{AB}} + c_2)}{(\mu_{\mathcal{A}}^2 + \mu_{\mathcal{B}}^2 + c_1)(\sigma_{\mathcal{A}}^2 + \sigma_{\mathcal{B}}^2 + c_2)}, \tag{3.4}$$

where

- $\mu_{\mathcal{A}}$ and $\mu_{\mathcal{B}}$ are, respectively, the average of pixel values in $\mathcal{A}$ and $\mathcal{B}$.
- $\sigma_{\mathcal{A}}$ and $\sigma_{\mathcal{B}}$ are, respectively, the variance of pixel values in $\mathcal{A}$ and $\mathcal{B}$.
- $\sigma_{\mathcal{AB}}$ is the covariance between pixel values of $\mathcal{A}$ and $\mathcal{B}$.
- $c_1 = (b_1\text{MAX})^2$ and $c_2 = (b_2\text{MAX})^2$ are two variables to stabilize the division with weak denominator. By default, $b_1 = 0.01$ and $b_2 = 0.03$.

Because bicubic interpolation serves as the baseline method, what is reported in the literature is ΔPSNR or ΔSSIM, i.e., the change made by an SR method over the bicubic interpolation baseline (as seen in Chapter 2).

As the foreground and background of EM images vary significantly, we also propose to segment the nanomaterial clusters (foreground) and the host material (background) through image binarization, and evaluate the improvements in PSNR and SSIM separately for the foreground as well as for the background. The foreground improvement reveals how well an SR method enhances the details of the image texture, whereas the background improvement points to a better de-noising capability.

The goal of super-resolution for EM images is to increase the ability of material characterization; for instance, increase the accuracy of morphology analysis. But PSNR and SSIM do not necessarily fully reflect a change in this capability. Thus we propose to add a metric to measure more directly the impact made by an SR method, which is to check whether the reconstructed images are able to facilitate a better detection of nanomaterial's boundary. For that, we use Canny's edge detector [44] to identify the boundaries and textures of the nanomaterial clusters and label the detected edges in a binary map. Let B_{HR} denote the binary edge map detected from the original HR image (ground truth) and B_{SR} denote the binary map detected from the reconstructed image resulting from the proposed SR method. The similarity between them is defined as:

$$\text{sim} = 1 - \frac{\|B_{\text{HR}} \neq B_{\text{SR}}\|_1}{\|B_{\text{HR}}\|_1 + \|B_{\text{SR}}\|_1}, \tag{3.5}$$

where $B_{\text{HR}} \neq B_{\text{SR}}$ produces an indicator matrix whose element is 1 where B_{HR} and B_{SR} have different values and 0 otherwise, and $\|\cdot\|_1$ is the entry-wise matrix 1-norm. A high sim indicates a better performance.

4. EXPERIMENTAL RESULTS*

4.1 General Results of PSNR and SSIM

With the two training options, self-training versus pooled-training, we test the following methods on the 22 pairs of SEM images:

- Two sparse-representation methods: the method proposed by Yang et al. [9] (2010) is abbreviated ScSR (Spare-coding based Super Resolution) and the method proposed by Trinh et al. [10] (2014) is abbreviated SRSW (Super-Resolution by Sparse Weight).
- Three deep learning-based super-resolution methods: VDSR (Kim et al. [11], 2016), EDSR (Lim et al. [12], 2017), and RCAN (Zhang et al. [13], 2018).
- Two versions of LB-NLM methods: the original LB-NLM (Sreehari et al. [16], 2017) and the paired LB-NLM (Qian et al. [14], 2019).

The 22 image pairs are partitioned into 198 in-sample subimages and 66 out-of-sample subimages. For ScSR, $L = 80,000$ paired patches of size 9×9 are randomly sampled to train a paired dictionary of size $1,024$. The same paired patches also make up the library for SRSW. VDSR, EDSR and RCAN are trained with their default settings and the data-augmentation and early-stopping options. For the original and paired LB-NLM methods, a paired library with the same size as in SRSW is constructed.

Table 4.1 presents the average improvement of PSNR and SSIM by these SR methods as compared with the bicubic interpolation baseline. We also report the percentages of the failure cases, which are defined as when an SR result yields a negative ΔPSNR.

The first observation is that for the paired image problem, self-training is a better strategy, despite the relatively small image sample size used. For all methods, self-training outperform

---

*© 2020 IEEE. Part of this chapter is reprinted with permission, from Y. Qian, J. Xu, L. F. Drummy and Y. Ding, "Effective Super-Resolution Methods for Paired Electron Microscopic Images," *IEEE Transactions on Image Processing*, vol. 29, pp. 7317-7330, 2020.

pooled-training in terms of the out-of-sample ΔPSNR. For most methods, the self-training also produces a better out-of-sample ΔSSIM while for some methods the pooled-training's ΔSSIM is better. But either way, the difference in ΔSSIM is marginal. As we argue earlier, using the learned relationship specific to a particular image pair pays off when that relationship is used for reconstruction. This pair-specific information does not exist in the general single-image SR when an external training set is used. Overall, self-training is indeed a better strategy because of its high accuracy and efficiency (training time is to be shown in Chapter 4.3).

Table 4.1: The improvements of PSNR and SSIM of the reconstructed SEM images after applying different SR methods, as compared with bicubic interpolation, and the percentages of the failure cases. The percentages of failure cases are also shown.

		Self-Training		Pooled-Training	
		In-Sample	Out-of-Sample	In-Sample	Out-of-Sample
ScSR [9]	ΔPSNR	0.26 dB	0.23 dB	0.18 dB	0.19 dB
	ΔSSIM	0.019	0.015	0.012	0.014
	Failure cases	7.2%		1.2%	
SRSW [10]	ΔPSNR	1.41 dB	1.17 dB	0.28 dB	0.31 dB
	ΔSSIM	0.026	0.026	0.019	0.022
	Failure cases	1.9%		16.3%	
VDSR [11]	ΔPSNR	2.22 dB	2.07 dB	1.24 dB	1.25 dB
	ΔSSIM	0.052	0.051	0.044	0.047
	Failure cases	0%		4.6%	
EDSR [12]	ΔPSNR	2.16 dB	2.06 dB	1.56 dB	1.35 dB
	ΔSSIM	0.052	0.052	0.050	0.051
	Failure cases	0%		4.5%	
RCAN [13]	ΔPSNR	2.24 dB	2.07 dB	1.84 dB	1.59 dB
	ΔSSIM	0.053	0.050	0.051	0.051
	Failure cases	0%		3.4%	
Original LB-NLM [16]	ΔPSNR	0.46 dB	0.45 dB	0.23 dB	0.28 dB
	ΔSSIM	0.016	0.016	0.017	0.018
	Failure cases	4.2%		10.6%	
Paired LB-NLM	ΔPSNR	3.75 dB	1.67 dB	0.87 dB	0.78 dB
	ΔSSIM	0.132	0.037	0.034	0.031
	Failure cases	0%		14.4%	

Among the methods in comparison, ScSR is not competitive when it is applied to the paired

EM images. The lack of competitiveness of ScSR can be explained by certain options used in its training process. ScSR extracts the high-frequency features from LR images. As the physical LR images contain heavy noisy, those high-frequency features do not adequately represent the image information. Also, ScSR assumes the reconstructed HR patches sharing the same mean and variance as the input LR patches, which is again not true for the physically captured image pairs. SRSW, on the other hand, obtains much better results by directly using the original patches. However, the randomly sampled library used in SRSW retains too many background patches with very little useful information. Such construction of the image library hampers SRSW's effectiveness. This shortcoming is to be further highlighted in the foreground/background analysis in Chapter 4.4.

Trained from the physically captured LR images, the performance of VDSR improves significantly as compared to the preliminary results in Chapter 2. In terms of both ΔPSNR and ΔSSIM, the three deep learning methods yield very similar results under self-training. Using pooled-training, the most advanced RCAN achieves the best performance but still is beaten by its self-training counterpart. A possible reason is that RCAN can benefit in pooled-training from its complex architectures, but this advantage, however, disappears in self-training. Considering the training time cost (to be shown in Chapter 4.3), VDSR under self-training appears to be the best candidate to the SR task for the paired EM images.

The simple, paired LB-NLM method achieves rather competitive performances and outperform the original LB-NLM, ScSR and SRSW. There are certain similarities between the paired LB-NLM method and SRSW. The paired LB-NLM method accounts for more factors behind the difference between a pair of physical images acquired at different resolutions, whereas SRSW primarily deals with the noise aspect. Both LB-NLM and SRSW show an obvious better performance when applied to the in-sample images under self-training, while for ScSR, the deep learning methods, and the original LB-NLM, the in-sample versus out-of-sample performance difference is much less pronounced.

The out-of-sample performance of the paired LB-NLM method under self-training reaches 80%

accuracy of the deep learning methods under the same setting. Considering the simplicity of the paired LB-NLM method, it is difficult to imagine that a simple method like that is able to achieve such a performance, relative to deep learning methods, on the general single-image SR problems; these results highlight the uniqueness of the SR problem for paired EM images.

In terms of the failure cases, the paired LB-NLM method and the three deep learning methods yield zero failure cases under self-training. Generally speaking, self-training produces fewer failure cases for all methods except ScSR than its pooled-training counterpart.

We present in Figures 4.1 and 4.2 the original LR images, bicubic interpolated images, the reconstructed images by the VDSR (both self-training and pooled-training), the reconstructed images by the paired LB-NLM method (self-training only), and the HR images (ground truth). Here VDSR is used as a representative of the three deep learning methods, since their respective best performances are similar. In each figure, four images are shown. The four images in Figure 4.1 are in-sample subimages, whereas those in Figure 4.2 are out-of-sample subimages. VDSR and the paired LB-NLM method both give us a clear foreground and a less noisy background. The visual results of the LB-NLM method are comparable to those of VDSR under self-training. The visual comparison between the images under VDSR (self-training) and those under VDSR (pooled-training) highlights the benefit of using the self-training strategy—the benefit of using self-training is particularly noticeable on the last two images, namely the last two rows of Figures 4.1 and 4.2.

Figures 4.1 and 4.2 also demonstrate the high noise level in both HR and LR EM images, which is a common phenomenon noted in the studies of EM images. The noise pattern is unique for each pair of EM images. For the paired image SR problems, the self-training strategy is more capable to learn the pattern and provide a more effective SR solution than pooled-training. It is worth noting that the background of the SR results from the three methods is clearer than that of the HR images. For the proposed LB-NLM method, we can explain it as a result of its inherent de-noising ability.

4.2 Comparison among Deep Learning Based Methods

As shown in Table 4.1, three deep learning based methods VDSR, EDSR and RCAN generate the largest improvement of PSNR and SSIM on out-of-sample subimages. Given this result, it is

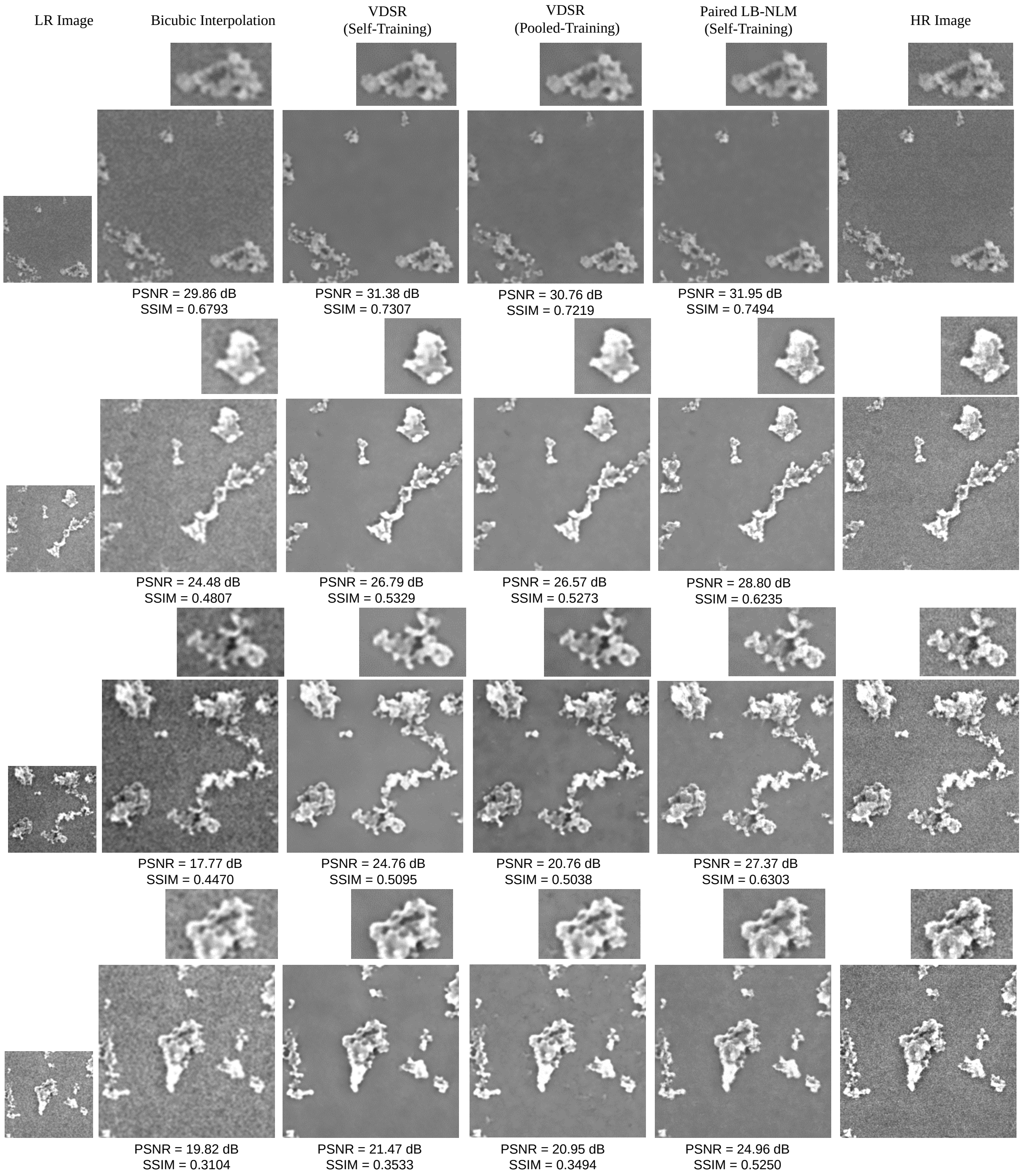


Figure 4.1: The LR images, the bicubic interpolated results, the image reconstruction results using VDSR (self-training and pooled-training), the paired LB-NLM method (self-training), and the ground truth (HR images) for four in-sample subimages. The small inserts in each row show a zoom-in of the foreground.

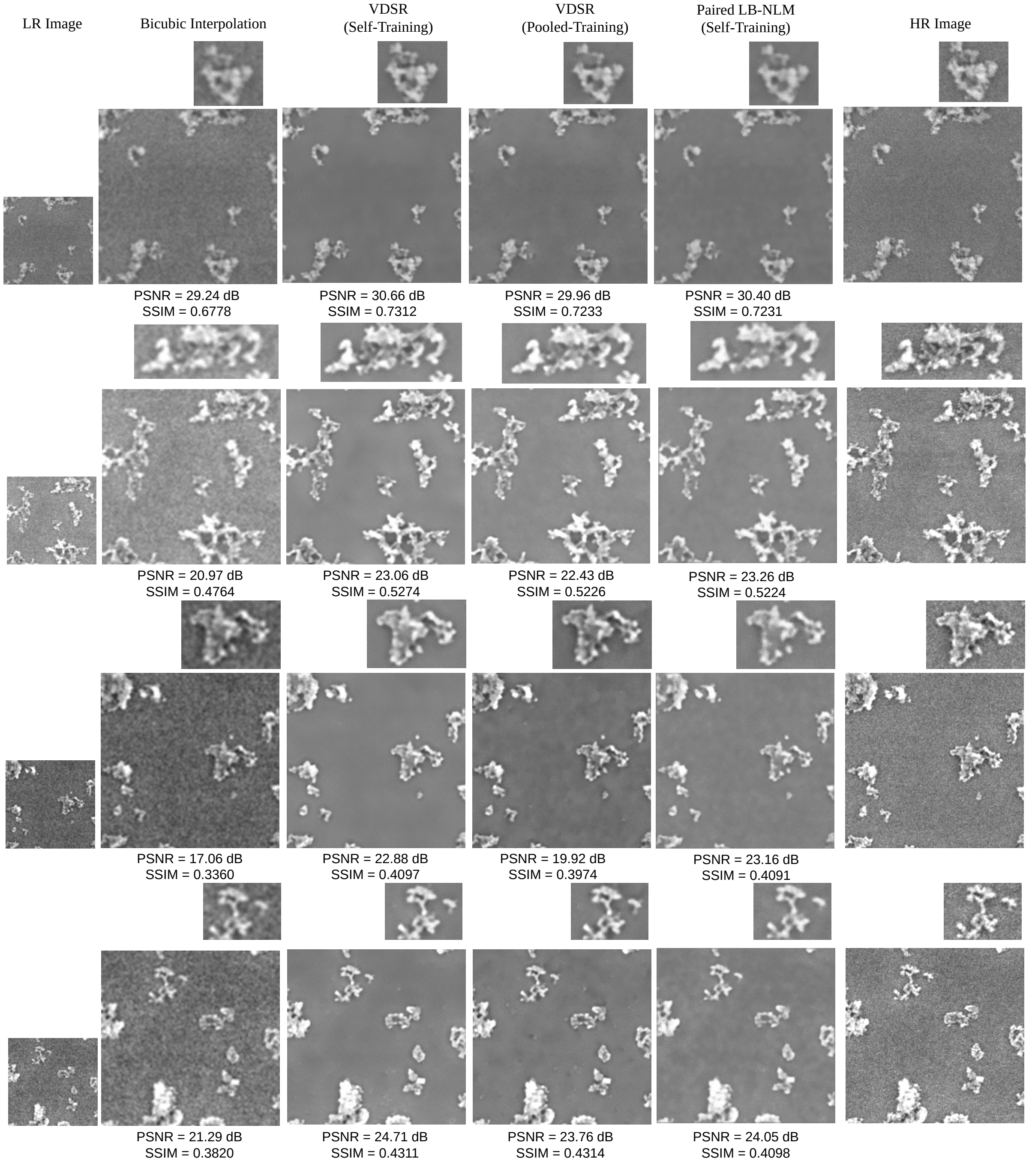


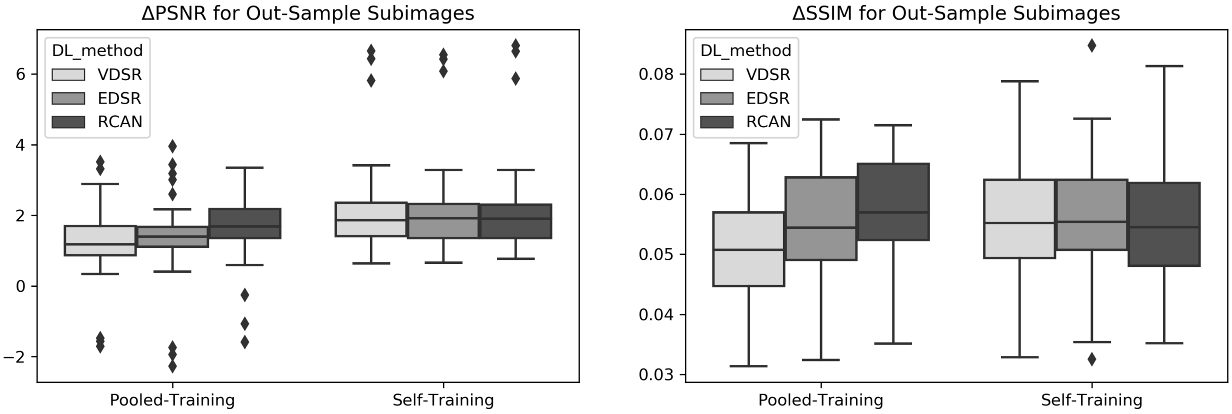
Figure 4.2: The LR images, the bicubic interpolated results, the image reconstruction results using VDSR (self-training and pooled-training), the paired LB-NLM method (self-training), and the ground truth (HR images) for four out-of-sample subimages. The small inserts in each row show a zoom-in of the foreground.

meaningful to dig into more details about the performance of these three methods. In this section, we focus on the out-of-sample ΔPSNR and ΔSSIM distribution of VDSR, EDSR and RCAN. With the three deep learning methods, we explore how two training strategies, i.e., self-training and pooled-training, affect each method.

About two training strategies, as we discussed before, self-training outperforms pooled-training in terms of ΔPSNR for all three deep learning methods, and in terms of ΔSSIM, self-training and pooled-training have comparable performance. This observation towards ΔPSNR counters the classic deep learning viewpoint, which is that more training data leads to a better performance. The training set of self-learning contains only 9 subimages. Compared with 198 subimages in pooled-training, it is a very small training set. In fact, even the amount of 198 images in a training set cannot compare with commonly used benchmark image datasets in computer vision, which normally contain thousands to millions of instances. This feature of relatively small size of training set results from the difficulty and high cost in getting such paired SEM images. However, in our experiment, the size of training set for self-training does not seem inadequate, as the super-resolution results show that using 9 pairs of training images, one can enhance the quality of the LR image substantially.

So why pooled-learning, which has much more training instances, cannot beat self-learning? The reason lies in discrepancies between different image pairs. Figure 4.3 presents the distribution of ΔPSNR and ΔSSIM for all 66 out-of-sample subimages, after applying three deep learning methods VDSR, EDSR and RCAN. We observe that for self-learning strategy, all out-of-sample subimages result in a positive ΔPSNR, and there are several rather substantial positive ΔPSNR around 6dB. In contrast, in the case of pooled-learning, there are several negative ΔPSNR values, while no positive ΔPSNR goes beyond 5dB. Since pooled-training aims to find a universal way to reconstruct all out-of-sample subimages, it has to fail at several subimages, which in turn hurts its overall performance.

From Figure 4.3, we cannot observe any major difference in the distribution of ΔSSIM between self-leaning and pool-training. VDSR's self-learning is a bit better than its pooled-training but the difference is not as much as ΔPSNR. All ΔSSIM range from 0.03 to 0.09 and no outlier. We conclude that two training strategies preserve the structural information of HR images equally well, when applying the deep learning networks.



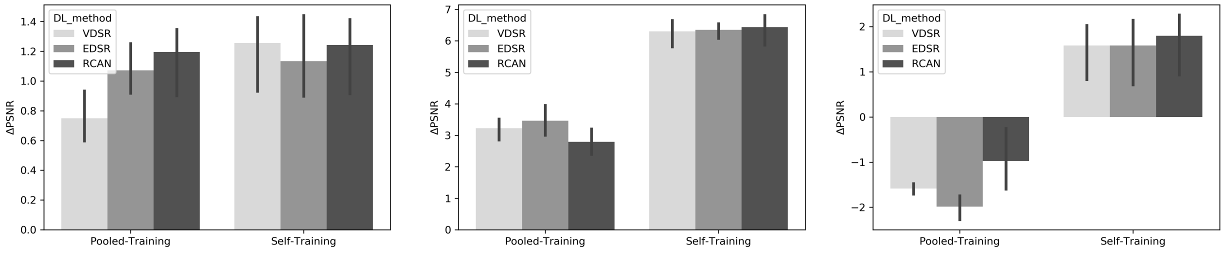
(a) (b)

Figure 4.3: The improvements of PSNR and SSIM distribution of all out-of-sample subimages after applying VDSR, EDSR and RCAN.

The most differentiating property of VDSR, EDSR and RCAN is their network architecture complexity. From the simplest to the most complicated, the order is VDSR, EDSR, and RCAN, which have, respectively, 20 layers, 69 layers, and 400 layers. Naturally, one may expect that the performance of VDSR, EDSR and RCAN also be in ascending order. Figure 4.3 shows that this viewpoint is true only if we consider the case of pooled-training. However, these deep learning methods under self-training do not exhibit noticeable differences, in terms of both ΔPSNR and ΔSSIM. It appears to us that when using self-training, VDSR performs equally well as RCAN, and is much less computationally expensive; see Table 4.2 in the next section when we present the computational times for a few SR methods.

When we look closer into the out-of-sample ΔPSNR made by the three deep learning methods over the total 22 image pairs, we find that the 22 image pairs can be classified into three groups.

Figure 4.4 shows the performance differences among the deep-learning methods in each group. The first group has 18 image pairs. In this group, the three deep learning methods VDSR, EDSR and RCAN have comparable performance under self-training, whereas under pooled-training, their performance does get better when the more complicated method is used. It is not surprising that such ΔPSNR pattern coincides with overall average performance in Table 4.1, because this group contains more than 80% of the image pairs. The second group has three image pairs. Its difference with the first group lies in the significant advantage of self-training over pooled-training. The rather substantial positive ΔPSNR values in Figure 4.3 come from this group. The third group contains only one pair. In this pair, self-training still works reasonably well (almost 2dB of ΔPSNR) but pooled-training fails to generate a positive ΔPSNR for all three deep networks. The last two groups make RCAN under pooled-training losing its edge over other networks under self-training. Apparently, the existence of the last two groups highlights the uniqueness of the paired images super-resolution problem, namely, for the paired images, one would think that a relationship learned directly from this specific pair is the best for boosting the resolution for the rest of the area uncovered by the high-resolution image.



(a) (b) (c)

Figure 4.4: Three patterns of improvement of PSNR distribution on out-of-sample subimages from the same pair of SEM images, where (a) is a representative of the first group (18 image pairs), (b) is a representative of the second group (3 image pairs), and (c) is group three (one image pair).

Our key finding regarding the three deep learning super-resolution methods, VDSR, EDSR, RCAN, with two training strategies, self-training and pooled-training, for paired SEM images can be summarized as:

- Currently the better training strategy for paired SEM images is self-learning, which can almost always produce a positive improvement of PSNR and SSIM for the reconstructed SEM images, as compared with bicubic interpolation, and its average improvement of PSNR is greater than pooled-training.

- Pooled-training results in a comparable ΔSSIM with self-training. Under pooled-training, more complex networks tend to perform better more often.

- A simple network such as VDSR under self-training strategy is so far the best super resolution method on paired SEM images. The disadvantage of self-training is that we need train a new network every time when we have a new pair of SEM images.

4.3 Computation Time

We present the computation time of training and inference for five methods: three deep learning methods, SRSW and the paired LB-NLM. We consider SRSW here as it is the better sparse-coding method. The three deep learning methods, implemented by PyTorch, are trained at Texas A&M University on one of its High Performance Research Computing (HPRC) Cluster with GPUs. The other two methods are trained with an MATLAB implementation on the same HPRC Cluster with parallel CPUs. Table 4.2 presents the average training and inference time when analyzing the 22 paired EM images.

With the aid of high computing power of GPUs, the deep learning methods still need a relatively long training time, especially when the pooled-training strategy is used. Training those models on a regular laptop computer without GPUs is not practical. Both SRSW and the paired LB-NLM methods can be trained efficiently and used on regular personal computers.

Concerning the inference time in Table 4.2, all deep learning methods run reasonably fast with GPUs, although RCAN is noticeably slower than the other two. We believe that the differences are caused by their network architectures. As mentioned in the previous section, VDSR uses 20 layers and EDSR uses 69 layers, but RCAN uses more than 400 layers. As RCAN adopts a much deeper network, its inference time becomes much longer. The paired LB-NLM's CPU time is similar to RCAN's GPU time. If all run on CPUs, LB-NLM is comparable or faster than the deep learning methods. SRSW suffers from a much longer inference time because it solves an L_1 optimization problem for each input LR patch.

As the self-training strategy produces the best results for the paired image SR problem, a user needs to re-train the model for every pair of newly captured images to attain the best enhancement. What this means is that unlike the traditional single-image SR problems, solving the paired SR problem prefers a method with a shorter training time. The long training time associated with the deep learning methods certainly puts them in a disadvantage. The proposed method, simpler and faster in training, presents itself as a competitive alternative, which can also be easily implemented on laptop computers without GPUs.

Table 4.2: Computation time of training and inference of some SR methods on HPRC Clusters.

	PyTorch (training on GPU and inference on GPU or CPU)			MATLAB (all on CPU)	
	VDSR [11]	EDSR [12]	RCAN [13]	SRSW [10]	Paired LB-NLM
Training Time	$\sim$ 30 mins (Self)	$\sim$ 30 mins (Self)	$\sim$ 10 hours (Self)	$\sim$ 5 mins (Both)	$\sim$ 5 mins (Both)
	$\sim$ 2 hours (Pooled)	$\sim$ 5 hours (Pooled)	$\sim$ 40 hours (Pooled)		
Inference Time	0.21 sec on GPU 4.08 sec on CPU	0.17 sec on GPU 2.40 sec on CPU	2.66 secs on GPU 21.65 sec on CPU	$\sim$ 25 mins	3.33 secs

4.4 Further Performance Analysis

In this section, we provide quantitative analysis using the new criteria for EM nanoimages: the separate foreground/background analysis and the edge detection analysis.

We first segment the SEM images by using Otsu's method [45] to highlight the separation of foreground from background and remove the isolated noise points in the foreground. Figure 4.5 shows the binary masks indicating the foreground versus the background in two images. Then we calculate separately the improvements of PSNR made by an SR method in the foreground and in the background.

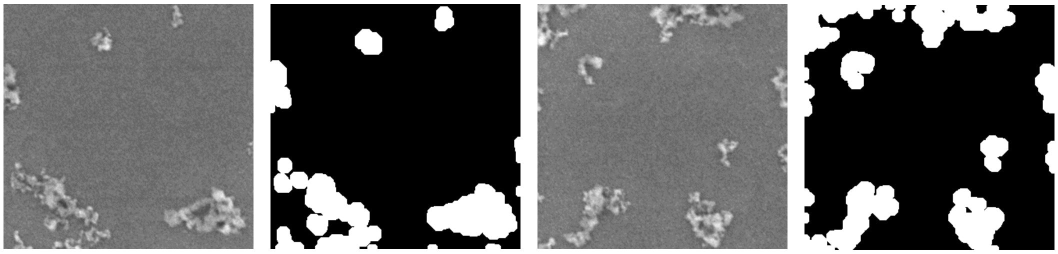

In-Sample Image Foreground Mask Out-of-Sample Image Foreground Mask

Figure 4.5: The foreground and background masks of an in-sample SEM subimage and an out-of-sample SEM subimage. The white areas indicate the nanomaterial (foreground), whereas the black areas indicate the host material (background).

Table 4.3: Changes in PSNR calculated for foreground and background for different SR results.

		VDSR [11]		SRSW [10]	Paired LB-NLM
		Self-Training	Pooled-Training	Self-Training	Self-Training
In-Sample	Foreground	1.21 dB	0.53 dB	0.03 dB	4.27 dB
	Background	2.86 dB	1.68 dB	2.27 dB	3.52 dB
Out-of-Sample	Foreground	0.97 dB	0.48 dB	−0.25 dB	0.23 dB
	Background	2.83 dB	1.75 dB	2.15 dB	2.65 dB

Table 4.3 presents the changes in PSNR for three methods: VDSR (both self-training and pooled-training), SRSW (self-training), and the paired LB-NLM method (self-training). It is apparent that all these methods denoise the background much more than they enhance the foreground. The main advantage of VDSR is its ability to improve the foreground better than the paired LB-NLM. This is not entirely surprising because the non-local-mean methods were originally designed as an image de-noising tool. It is also observed that the self-training VDSR is better than the pooled-training VDSR more so in terms of a stronger de-nosing capability over the background. SRSW does a similar job in terms of denoising the background. But there is a slight decrease in terms of PSNR for the foreground, which suggests that the particular mechanism used in SRSW, especially the mechanism to create its library, is not effective for enhancing the foreground signals in the physical EM images.

Next we apply Canny's edge detector [44] to the HR images, the bicubic interpolated images,

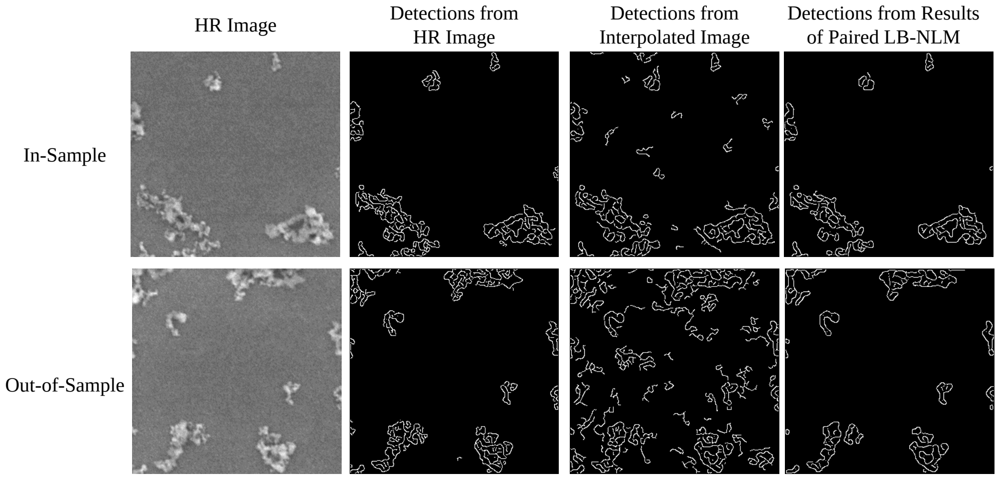


Figure 4.6: The results of Canny edge detection from the HR images and some reconstructed images.

Table 4.4: Results of sim for different SR methods and bicubic interpolation by Canny's detector.

	Bicubic Interpolation	VDSR [11]		SRSW [10]	Paired LB-NLM
		Self-Training	Pooled-Training	Self-Training	Self-Training
In-Sample	0.25	0.39	0.37	0.33	0.56
Out-of-Sample	0.24	0.37	0.35	0.24	0.33

and the reconstruction images by the three methods mentioned above. A key parameter in Canny's edge detector is set as 0.2. Figure 4.6 demonstrates the detection results. The visual inspection show that the SR reconstructed results facilitate better edge detection than the bicubic interpolation baseline method.

To quantify the improvement in detection accuracy, we in Table 4.4 calculate the similarity index, sim, as defined in equation (3.5) in Chapter 3.3. Except for SRSW, all methods can improve, as compared with the bicubic interpolation baseline, the Canny's detection accuracy by around 50% on the out-of-sample images. The self-training VDSR achieves the largest improvement, although its sim is just slightly higher than that of the pooled-training VDSR and the paired LB-NLM. These results is consistent with the foreground PSNR improvements made by the four method in Table 4.3.

5. CONCLUSIONS AND FUTURE WORK*

We present in this paper the paired EM image super-resolution problem and report our investigation of how best to address this problem.

Paired images are not very common in public image databases because taking them needs special care and specific setup. On the other hand, they are rather common in scientific experiments, especially in material and medical research. The use of electron microscopes exacerbates the need for handling paired images for the purpose of super-resolution. Unlike optical photographing, the imaging process using an electron microscope is not non-destructive. Imaging using high-energy electron beams can damage sample specimen and must be carefully administrated. Consequently, researchers tend to use low-energy beams or subject the samples to a short duration of exposure. The results are of course low-resolution images. An effective super-resolution method that can subsequently boost these low-resolution images to a higher resolution has a significant impact on scientific fields relying on electron imaging.

In this research, we compare different state-of-the-art super-resolution approaches for handling the paired EM image problem. The take-home messages of our research can be summarized as follows:

- For the paired image problem, a local registration is important, as it accounts for the distortion between the image pairs. Our current approach is adequate but there is an ample room for further improvement.
- When presented with paired images, it is recommended to use the self-training strategy, in spite of the relatively small sample size under that circumstance. Under self-training, simpler SR solutions are demonstrably as effective as more complex models.
- Deep learning SR methods exhibit their extraordinary performances on paired EM images.

When using self-training strategy, a simpler network structure is preferred, whereas a deeper network show better performance under pooled-training. The paired EM image SR problem has its unique performance distribution due to the intrinsic properties of paired EM images.

- The paired LB-NLM entertains the advantage of fast training and better interpretability and is a close runner-up to the deep learning methods. It can be readily implemented on ordinary laptop computers.

Our work is among the very early efforts in addressing the paired EM image super-resolution problem. We see two broad future research directions as moving forward: (a) It is worthwhile to explore a deep neural network with a specialized architecture designed for the paired EM image problems. When the deep network can account for the two uniqueness in the problem (i.e., the image pairing and the electron images), a much greater enhancement of the low-resolution images can be anticipated; (b) It is interesting to observe the competitiveness of the simple, paired LB-NLM method or the simpler deep-learning network like VDSR. By exploiting the property and uniqueness of the paired image problems, it is possible to develop a computationally simple and structurally more interpretable method with good effectiveness.

REFERENCES

[1] C. Park, J. Huang, D. Huitink, S. Kundu, B. Mallick, H. Liang, and Y. Ding, "A multi-stage, semi-automated procedure for analyzing the morphology of nanoparticles," *IIE Transactions Special Issue on Nanomanufacturing*, vol. 44, no. 7, pp. 507–522, 2012.

[2] C. Park, J. Huang, J. Ji, and Y. Ding, "Segmentation, inference and classification of partially overlapping nanoparticles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 669–681, 2013.

[3] H. Yang and N. Ahuja, "Automatic segmentation of granular objects in images: Combining local density clustering and gradient-barrier watershed," *Pattern Recognition*, vol. 47, no. 6, pp. 2266–2279, 2014.

[4] Y. Qian, J. Z. Huang, X. Li, and Y. Ding, "Robust nanoparticles detection from noisy background by fusing complementary image information," *IEEE Transactions on Image Processing*, vol. 25, no. 12, pp. 5713–5726, 2016.

[5] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine*, vol. 20, no. 3, pp. 21–36, 2003.

[6] J. Tian and K.-K. Ma, "A survey on super-resolution imaging," *Signal, Image and Video Processing*, vol. 5, no. 3, pp. 329–342, 2011.

[7] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Processing*, vol. 128, no. 11, pp. 389–408, 2016.

[8] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, July 21–26 2017, pp. 1110–1121.

[9] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[10] D.-H. Trinh, M. Luong, F. Dibos, J.-M. Rocchisani, C.-D. Pham, and T. Q. Nguyen, "Novel

example-based method for super-resolution and denoising of medical images," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1882–1895, 2014.

[11] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, June 27–30 2016, pp. 1646–1654.

[12] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, July 21–26 2017, pp. 136–144.

[13] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Munich, Germany, September 8–14 2018, pp. 286–301.

[14] Y. Qian, J. Xu, L. F. Drummy, and Y. Ding, "Effective super-resolution methods for paired electron microscopic images," *IEEE Transactions on Image Processing*, vol. 29, pp. 7317–7330, 2020.

[15] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, San Diego, CA, USA, June 20–26 2005, pp. 60–65.

[16] S. Sreehari, S. Venkatakrishnan, K. L. Bouman, J. P. Simmons, L. F. Drummy, and C. A. Bouman, "Multi-resolution data fusion for super-resolution electron microscopy," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Honolulu, HI, USA, July 21–26 2017, pp. 1084–1092.

[17] T. Komatsu, T. Igarashi, K. Aizawa, and T. Saito, "Very high resolution imaging scheme with multiple different-aperture cameras," *Signal Processing: Image Communication*, vol. 5, no. 5-6, pp. 511–526, 1993.

[18] N. R. Shah and A. Zakhor, "Resolution enhancement of color video sequences," *IEEE Transactions on Image Processing*, vol. 8, no. 6, pp. 879–885, 1999.

[19] W. T. Freeman, T. R. Jones, and E. C. Pasztor, “Example-based super-resolution,” *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[20] H. Chang, D.-Y. Yeung, and Y. Xiong, “Super-resolution through neighbor embedding,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Washington, DC, USA, June 27 – July 2 2004, pp. 1–8.

[21] T.-M. Chan, J. Zhang, J. Pu, and H. Huang, “Neighbor embedding based super-resolution algorithm through edge detection and feature selection,” *Pattern Recognition Letters*, vol. 30, no. 5, pp. 494–502, 2009.

[22] X. Gao, K. Zhang, D. Tao, and X. Li, “Image super-resolution with sparse neighbor embedding,” *IEEE Transactions on Image Processing*, vol. 21, no. 7, pp. 3194–3205, 2012.

[23] P. Sandeep and T. Jacob, “Single image super-resolution using a joint GMM method,” *IEEE Transactions on Image Processing*, vol. 25, no. 9, pp. 4233–4244, 2016.

[24] Y. Huang, J. Li, X. Gao, L. He, and W. Lu, “Single image super-resolution via multiple mixture prior models,” *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 5904–5917, 2018.

[25] K. S. Ni and T. Q. Nguyen, “Image superresolution using support vector regression,” *IEEE Transactions on Image Processing*, vol. 16, no. 6, pp. 1596–1610, 2007.

[26] K. I. Kim and Y. Kwon, “Single-image super-resolution using sparse regression and natural image prior,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 1127–1133, 2010.

[27] R. Timofte, V. De Smet, and L. Van Gool, “A+: Adjusted anchored neighborhood regression for fast super-resolution,” in *Proceedings of the Asian Conference on Computer Vision (ACCV)*, Singapore, Singapore, November 1–5 2014, pp. 111–126.

[28] H. Wang, X. Gao, K. Zhang, and J. Li, “Single-image super-resolution using active-sampling gaussian process regression,” *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 935–948, 2016.

[29] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, “Coupled dictionary training for image

super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467–3478, 2012.

[30] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, RI, USA, June 18–20 2012, pp. 2216–2223.

[31] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, Zurich, Switzerland, September 6–12 2014, pp. 184–199.

[32] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, June 18–23 2018, pp. 2472–2481.

[33] L. Xu, J. S. Ren, C. Liu, and J. Jia, "Deep convolutional neural network for image deconvolution," in *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*, Montreal, Canada, December 8–13 2014, pp. 1790–1798.

[34] C. J. Schuler, M. Hirsch, S. Harmeling, and B. Schölkopf, "Learning to deblur," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 7, pp. 1439–1451, 2015.

[35] X. Xu, J. Pan, Y.-J. Zhang, and M.-H. Yang, "Motion blur kernel estimation via deep learning," *IEEE Transactions on Image Processing*, vol. 27, no. 1, pp. 194–205, 2017.

[36] R. Timofte, S. Gu, J. Wu, and L. Van Gool, "NTIRE 2018 challenge on single image super-resolution: Methods and results," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Salt Lake City, UT, USA, June 18–23 2018, pp. 852–863.

[37] M. Grubinger, P. Clough, H. Müller, and T. Deselaers, "The IAPR TC-12 benchmark: A new evaluation resource for visual information systems," in *Proceedings of the OntoImage Workshop on Language Resources for Content-based Image Retrieval*, Genoa, Italy., May 22 2006, pp. 13–22.

[38] R. Keys, “Cubic convolution interpolation for digital image processing,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[39] X. Xu, D. Sun, J. Pan, Y. Zhang, H. Pfister, and M.-H. Yang, “Learning to super-resolve blurry face and text images,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, October 16–20 2017, pp. 251–260.

[40] X. Xu, Y. Ma, and W. Sun, “Towards real scene super-resolution with raw images,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June 16–20 2019, pp. 1723–1731.

[41] X. Zhang, Q. Chen, R. Ng, and V. Koltun, “Zoom to learn, learn to zoom,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, June 16–20 2019, pp. 3762–3770.

[42] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, “Distributed optimization and statistical learning via the alternating direction method of multipliers,” *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[43] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[44] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–698, 1986.

[45] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.