A Super-Resolution Training Paradigm Based on Low-Resolution Data Only to Surpass the Technical Limits of STEM and STM Microscopy

Björn Möller[1] Jan Pirklbauer[1] Marvin Klingner[1] Peer Kasten[2]
Markus Etzkorn[2] Tim J. Seifert[2] Uta Schlickum [2] Tim Fingscheidt[1]
{bjoern.moeller, j.pirklbauer, m.klingner, p.kasten,
m.etzkorn, johannes.seifert, u.schlickum, t.fingscheidt}@tu-bs.de
Technische Universität Braunschweig, Germany [1]Institute for Communications Technology [2]Institute of Applied Physics

Abstract

Modern microscopes can image at atomic resolutions but often reach technical limitations for high-resolution images captured at the smallest nanoscale. Prior works have applied super-resolution (SR) by deep neural networks employing high-resolution images as targets in supervised training. However, in practice, it may be impossible to obtain these high-resolution images at the smallest atomic scales. Approaching this problem, we consider a new super-resolution training paradigm based on low-resolution (LR) microscope images only, to surpass the highest physically captured resolution available for training. As a solution, we propose a novel multi-scale training method for SR based on LR data only, which simultaneously supervises SR at multiple resolutions, allowing the SR to generalize beyond the LR training data. We physically captured low- and high-resolution images for evaluation, thereby incorporating real microscope degradation to deliver a proof of concept. Our experiments on periodic atomic structure in STEM and STM microscopy images show that our proposed multi-scale training method enables deep neural network image SR even up to 360% of the highest physically recorded resolution. Code and data is available on github[1].

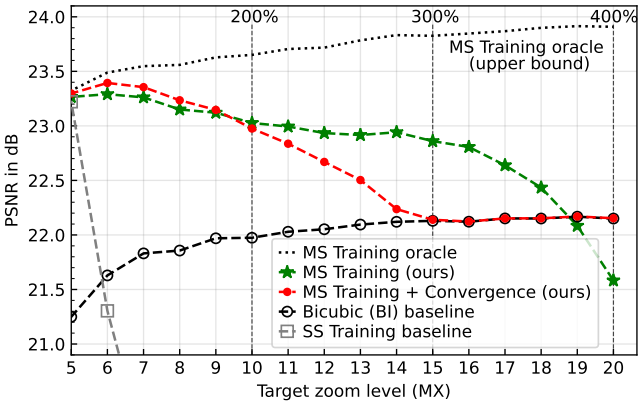


Figure 1. **STEM PSNR** results with **standard evaluation** of a **4x SR** with our proposed multi-scale (MS) training vs. a single-scale (SS) and a bicubic interpolation (BI) baseline, depending on the target zoom level with millionfold magnification (MX) of microscope images of **gallium nitride** (GaN). MS and SS models were trained based on low-resolution 5MX STEM images $\mathcal{D}^{\text{train}}_{\text{STEM}-5\text{MX}}$, the MS training oracle based on (downsampled) high-resolution 20MX STEM images. All methods were evaluated on high-resolution 20MX STEM images $\mathcal{D}^{\text{test}}_{\text{STEM}-20\text{MX}}$ and pseudo-real downsampled versions thereof (5MX...19MX). For target zoom levels above 5MX, the proposed MS training leads to better image quality than the SS and the BI baselines.

1. Introduction

Single image super-resolution (SR) describes the task of creating a higher resolution and higher quality image from a given typically physically captured low-resolution image. Deep learning based SR techniques have been adopted to various microscopy applications, such as medicine [1], cell biology [16] and physics [10]. One promising application are scanning microscopes, such as scanning transmission electron microscopes (STEM) [38] and scanning tunneling microscopes (STM) [3,37], which allow to access the structure and properties of solid state matter down to the atomic scale. They raster the sample using a probe and produce 2D images. However, capturing images at a higher pixel density at a certain limit comes at the cost of decreasing image quality due to time dependent drift effects [38], which practically limits maximum pixel resolution. As these microscopes capture images point by point, causing the acquisition time to scale quadratically with image resolution, the effect of drift also becomes more prominent. While some of these noise effects can be minimized by more expensive equipment, allowing measurements at low temperature, the general problem of drift effects is always present.

[1]https://github.com/ifnspaml/SuperResolutionMultiscaleTraining

To overcome these limitations, DNN super-resolution [9, 11, 22–24, 27, 32, 39, 40] is investigated as an intriguing approach to super-resolve images beyond the capabilities of those scanning microscopes. SR deep neural networks (DNNs) are usually trained in a supervised manner, meaning that the reconstruction of a high-resolution (HR) image from a low-resolution (LR) input is learned, requiring paired LR/HR images [9, 11, 22–24, 27, 32, 39, 40]. For scanning microscopes operating at nanoscale, it might not be possible to measure high-resolution (HR) high-quality reference images to learn the reconstruction from.

Motivated by these technical limitations, we consider a new super-resolution training paradigm based on low-resolution (LR) microscope images only. Following this paradigm, our proposed multi-scale (MS) training method generates new LR/HR training pairs from the available LR images of a certain single zoom level, enabling training in a supervised fashion. The approach is based on a configurable data augmentation combining various image interpolation techniques. MS training utilizes upscaling of the available LR images to create multiple resolutions to crop new HR training targets from, and then degrades these to obtain LR training inputs. While image SR increases pixel resolution, we use the term *zoom level* to describe the amount of pixels that are used within an image to depict a given fixed area on a sample, thereby incorporating magnification.

Fig. 1 shows that MS training (green curve) enables DNN SR based on LR images only, significantly outperforming a common single-scale (SS) training. As SR performance decreases for target zoom levels far beyond that of the LR data, the method can be configured to converge to bicubic interpolation for high target resolutions (red curve).

For our experiments, we captured data at two magnifications, such that physically recorded images of a low and of a high zoom level are available, both of which contain some degree of pixel-based noise. To quantify the methods' results, we use the high zoom level images as references during evaluation, while training is performed on low zoom level images only. Following supervised SR works [9, 11, 22–24, 27, 32, 39, 40], we do a *standard evaluation* on degraded versions of the reference images, although using a noise-preserving degradation. Additionally, we present a *physical evaluation*, which uses physically recorded lower zoom level images (LR) as model inputs and physically recorded higher zoom level images (HR) as references, incorporating real-world microscope image degradation into our evaluation.

Our contributions are the following. First, we propose a multi-scale training method for SR models, which we apply to periodic atomic structure STM and STEM microscopy images. Second, we show that our method generalizes to resolutions beyond those available in the training data. Third, we provide a proof of concept for image SR beyond the technical limits of microscopes, potentially increasing the zoom levels of microscopes around the world.

2. Related Work

Image super-resolution Traditionally, convolutional neural networks (CNNs) are well suited for SR [11, 12, 20, 24, 31, 40], while newer techniques use generative adversarial networks (GANs) [21] and with the recent surge of attention [33] for vision models [13], transformer-based architectures have shown state-of-the-art performance for this task [7, 23, 26]. In this work, we chose the `SwinIR` transformer model [23] for our image SR experiments. These models were trained in a supervised fashion, reconstructing HR reference images from LR-degraded versions thereof, often obtained by a simple degradation simulation, i.e., the direct downscaling using the bicubic kernel [6]. Only few authors collect both resolutions physically [5, 10], since the process is tedious due to imperfections in the image acquisition. In this work, we collect both resolutions but assume that HR reference images are not available for training supervision but just during evaluation.

Few initial works tackle this unsupervised SR problem [2, 30, 34]. Shocher et al. [30] proposed a zero-shot SR method (ZSSR), training a CNN for each LR test image on generated LR/HR training pairs, downscaled, cropped and degraded from the LR image. Building on that, Ahn et al. [2] train a model using multiple LR images. Another approach utilizes GANs to learn super-resolution and downscaling simultaneously from LR images [34]. However, these methods struggle with noisy real-world LR images. Both [2, 30] rely on bicubic kernels for image degradation, filtering pixel noise in this process, while Wang et al. [34] assume a zero-noise scenario. However, simple scanning microscopes produce rather noisy images, making these approaches appear suboptimal. Accordingly, our multi-scale training integrates a dedicated noise-preserving downscaling to simulate real degradation in input images. Furthermore, our method generates LR/HR training pairs by not only downscaling the LR image but also by upscaling it, thereby creating HR targets at multiple scales for training. Per HR target, we also create multiple degraded versions as training inputs via a set of various interpolation functions, including a noise-preserving degradation.

Super-resolution for scanning microscopy images Recent works demonstrate that image reconstruction techniques are able to generate high-resolution microscopy images from low-resolution scanning microscope measurements. They especially focus on the increased acquisition speed by utilizing sparse scanning patterns in combination with learnable image reconstruction [15, 19]. To generate the LR/HR training pair for supervised training, some approaches mask out pixels in the HR image to yield a sparse

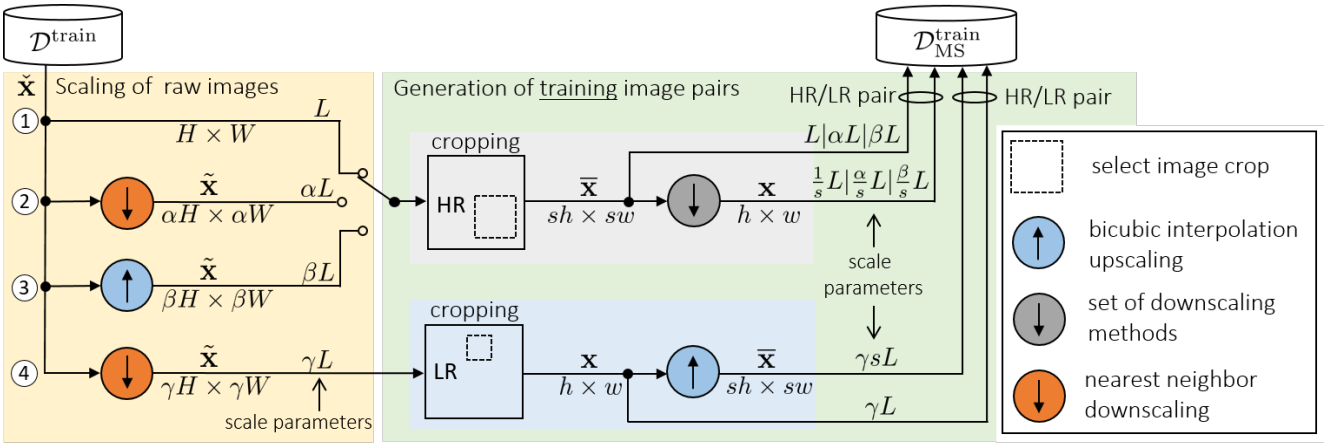


Figure 2. Overview of our **proposed multi-scale training method (data generation)** using a combination of image interpolation functions to enable super-resolution beyond the highest physically captured zoom level available in the training data. Training data, consisting of HR/LR pairs of image crops ($\overline{\mathbf{x}}$, $\mathbf{x}$), is generated that spans multiple zoom level scales.

LR image [15], while others register physically captured images of the same sample as LR and HR training pairs [10,16,25]. In contrast to our work, these approaches follow the supervised training paradigm having access to the HR images to learn the reconstruction from, whereas we perform training only on the basis of LR measurements. Also, hardly any of the former approaches considers microscopy at atomic scale [15], which we operate on. Finally, we physically capture images at a high and low resolution [16], and show how to evaluate on unpaired LR/HR images.

3. Proposed Super-Resolution Training

3.1. Training Paradigm

Deep learning based SR models are trained to reconstruct a higher-resolution (HR) image from a given lower-resolution (LR) image. In practice, however, with scanning microscopes, noise effects become more prominent as magnification increases, making it difficult or even impossible to capture high-quality images at the smallest nanoscales. Therefore, we assume that HR images cannot be acquired for training, but only microscope data at low resolution. Approaching this problem, we consider a new training paradigm for training super-resolution models based on LR microscope images only. To the best of our knowledge, we are the first to address the LR-only training problem for microscope data, and for noisy LR training data in general.

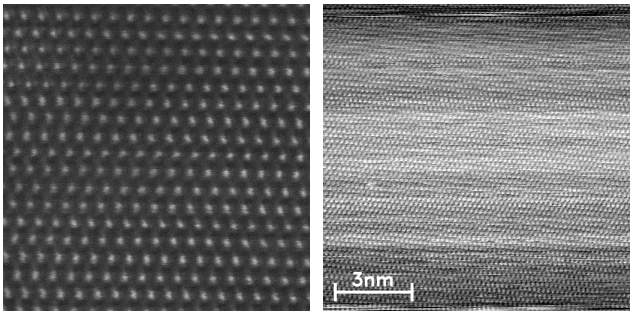
3.2. Multi-Scale (MS) Training Method

Our proposed approach is a multi-scale training method, trainable from LR data only, enabling supervised single-image super-resolution models to deliver microscope images *beyond the maximum of physically captured zoom levels in the training data*. The approach is based on data augmentation by a combination of various image interpolation techniques for image resizing. The method iterates over a training dataset $\mathcal{D}^{\text{train}}$ and takes a raw grayscale image $\check{\mathbf{x}} \in \mathbb{G}^{H \times W}$ as input, where $\mathbb{G} = [0, 255]$ is the set of gray values, while H, W define height and width, respectively.

Fig. 2 provides an overview of the core of our proposed multi-scale training method, which is the dataset generation for training. Since the data is prepared for supervised training, the data generation's output is a set of paired high-resolution (HR) target images $\overline{\mathbf{x}}$ of size $sh \times sw$ and low-resolution (LR) input images $\mathbf{x}$ of size $h \times w$, whose resolutions differ by the super-resolution factor $s > 1$. Each training pair is generated by one out of four unique augmentation sub-processes (① - ④). In sub-process ①, for a raw low-resolution image $\check{\mathbf{x}}$, HR training targets $\overline{\mathbf{x}}$ are simply obtained by cropping with no image scaling involved. These crops are degraded to LR training model inputs using a set of various downscaling methods (①). The set of image scaling functions consists of nearest neighbor [4], Lanczos [14], bilinear [4], bicubic [4], box [17] and Hamming interpolation [18], resulting in six training pairs per crop. In contrast to traditional image SR training [9, 11, 22–24, 27, 32, 39, 40], our proposed multi-scale training aims to model a variety of HR↔LR transformations by additionally scaling the source images up, applying bicubic interpolation (③), and down, applying nearest neighbor interpolation (②, ④), thereby obtaining images *both* at higher and lower resolution. In any case ①, ②, ③, ④, crops are generated. For ①, ②, ③, crops are of size $sh \times sw$, while for ④, crops are of size $h \times w$. For LR model input generation, ② and ③ further follow the degradation process of ①. Sub-process ④, however, takes the crops as LR input images $\mathbf{x}$, and generates the respective HR target image $\overline{\mathbf{x}}$ by bicubic upscaling. The method's sub-processes ②, ③, and ④ can be applied multiple times with different values for factors $\alpha \in \mathcal{A}$, $\beta \in \mathcal{B}$ and $\gamma \in \Gamma$, resulting in

Table 1. **Multi-scale training configurations.** The parameters α, β and γ control the target zoom level ranges of the multi-scale training data generated by the method's sub-processes (see Fig. 2, ②-④). Resulting target zoom level ranges (MX, nm) are given in the far right columns. Markers refer to sub-process configurations as used in the result figures (see Figs. 1, 5, 6, 7, 8).

	SR	Marker	Parameters			Target ranges	
			α	β	γ	①,②,③	④
STEM	4x (MX)	▼	0.25:0.95	-	-	1.25-5	-
		▲	-	1.1:2	-	5-10	-
		★	0.25:0.95	1.1:2	-	1.25-10	-
		◆	0.25:0.95	1.1:2	0.55:0.75	1.25-10	11-15
		◆	0.25:0.95	1.1:3.4	0.85:0.95	1.25-17	17-19
		●	-	1.1:3.4	0.85:0.95	5-17	17-19
STM	2x (nm)	▲	-	1.04:1.71	-	24-14	-
		★	0.5:0.96	1.04:1.71	-	48-14	-
		●	-	1.04:1.71	0.86:0.92	24-14	14-13



(a) GaN sample (b) Graphite surface

Figure 3. Physically captured microscopy images: (a) **STEM** image of a gallium nitride (GaN) sample with magnification of 20MX (b) **STM** image of a graphite surface with magnification to 12nm.

a multi-scale zoom level range. Since β specifies a factor for upscaling, we have $\mathcal{B} \subset (1, \infty)$ (③), while $\mathcal{A} \subset (0, 1)$, $\Gamma \subset (0, 1]$ represent ranges of downscaling factors (②, ④).

We use various multi-scale training configurations generating training data based on LR STEM data $\mathcal{D}^{\text{train}}_{\text{STEM}-5\text{MX}}$ and LR STM data $\mathcal{D}^{\text{train}}_{\text{STM}-24\text{nm}}$, the specifics of which we show in Table 1. It displays the parameterization of α, β, and γ and corresponding notations incorporating the MX or nm scale for better readability of results. The step size within the respective intervals $\alpha_{\min} : \alpha_{\max}$, $\beta_{\min} : \beta_{\max}$, and $\gamma_{\min} : \gamma_{\max}$ corresponds to integer MX or nm values (see "target ranges" column), with the exception of $\alpha_{\min} = 0.25$ and $\alpha_{\max} = 0.95$, where a step size of 0.05 for 4xSR is used. As notation in the later results figures, we list the sub-processes used in the respective configuration and, following in parentheses, the generated range of target zoom levels (MX, nm) from minimum to maximum. The ranges for ①,②,③ are listed jointly, since they use the same degradation process to generate LR images. For example, "MS T.①,②,③(1.25-17) ④(17-19)" denotes a multi-scale (MS) training, where the first three sub-processes operate with target zoom levels ranging between the minimum and maximum of 1.25 and 17. Training data was also generated by ④ for target zoom levels 17, 18, 19. "MS T.①,③(5-10)" indicates the integer target zoom level range 5...10 and that ②, ④ were not used for training data generation (no α, γ).

Table 2. **Microscopy datasets** for **STEM** and **STM**: Images showing periodic structures of solid matter are investigated. "Scale" refers to magnification (STEM) or is the nanoscale image size (STM). "Scale parameter" L is proportional to the STEM (MX) scale and antiproportional to the STM (nm) scale.

	Scale	L	# Samples in train/val/test			Notation	
STEM	5MX	1	8	1	1	$\mathcal{D}_{\text{STEM}-5\text{MX}}$	$= \mathcal{D}_{\text{L}_{\text{train}}}$
	20MX	4	7	1	1	$\mathcal{D}_{\text{STEM}-20\text{MX}}$	$= \mathcal{D}_{\text{sL}_{\text{train}}}$
STM	24nm	1	205	21	21	$\mathcal{D}_{\text{STM}-24\text{nm}}$	$= \mathcal{D}_{\text{L}_{\text{train}}}$
	12nm	2	439	45	45	$\mathcal{D}_{\text{STM}-12\text{nm}}$	$= \mathcal{D}_{\text{sL}_{\text{train}}}$

4. Datasets and Evaluation Setup

4.1. Microscope Datasets

In this work, we either use real data recorded by a scanning transmission electron microscope (STEM) or by a scanning tunneling microscope (STM). All samples from the STM show graphite surfaces, while the STEM images show gallium nitride (GaN). The STEM was a `JEOL Neoarm F200` with aberration correction, while the STM images have been captured by a table-top `Nanosurf NaioSTM` at room temperature under air. Example images are shown in Fig. 3. All images depict materials with periodic atomic structure, which we focus on. Table 2 provides an overview of the datasets used. The scale refers to a magnification (STEM) or is the nanometer image width and height (STM). We introduce the *scale parameter* L, which is proportional to the STEM (MX) magnification and antiproportional to the STM (nm) scale, thereby harmonizing both the STEM and the STM scales. A large L indicates a higher zoom level. The number of samples refers to images of size 2048x2048 for STEM and 512x512 for STM. For the experiments, we split the data into training, validation, and test sets proportionately at about 80%, 10%, and 10%, respectively.

4.2. Overall Evaluation Setup

To evaluate our method, we use the peak signal-to-noise ratio (PSNR) [29] and the multi-scale structural similarity index measure (MS-SSIM) [36] as common image quality metrics, which are established in the fields of image reconstruction and SR. For evaluation, we have access to two datasets captured at magnification levels differing by super-resolution factor s. To still evaluate the methods for zoom levels between those, we also generate downscaled realistic versions of the HR reference images by downscaling with

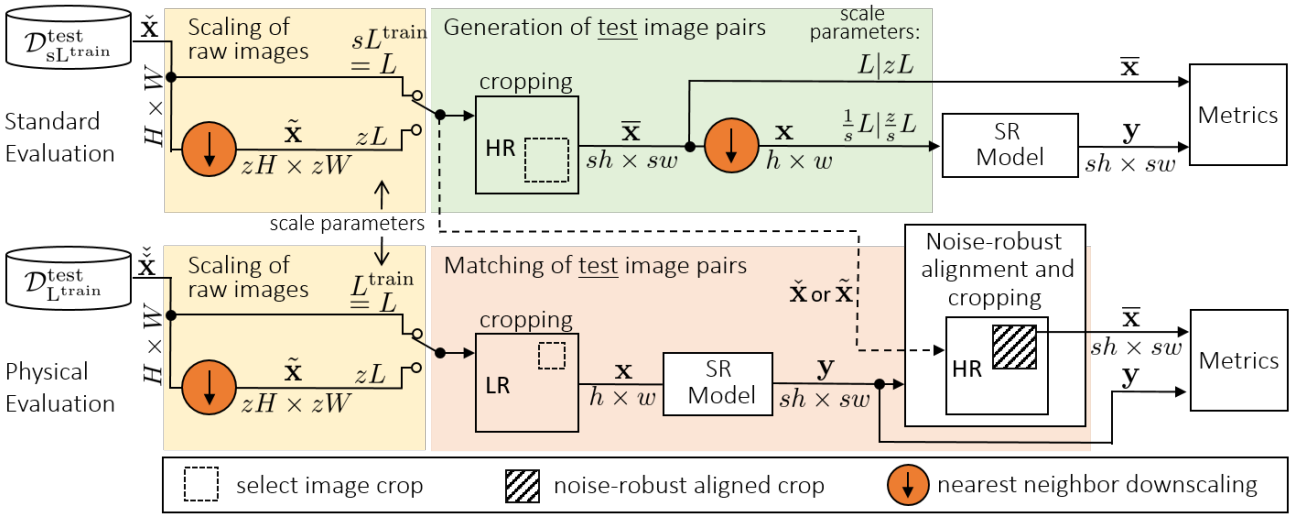


Figure 4. Overview of our **evaluation setup**. Image quality metrics are calculated from the SR model's HR prediction $\mathbf{y}$ and a HR target image $\overline{\mathbf{x}}$. **Standard evaluation** (top) uses crops $\overline{\mathbf{x}}$ of images taken from a test set with s times the training scale parameter $\mathrm{sL}^{\mathrm{train}} = L$ or downsampled versions thereof as HR targets *and degrades them* to yield LR SR model inputs $\mathbf{x}$. **Physical evaluation** (bottom) uses crops of test images with training scale parameter $\mathrm{L}^{\mathrm{train}}$ as LR inputs and matches the model's prediction $\mathbf{y}$ against images *from a different dataset with a higher scale parameter* $\mathrm{sL}^{\mathrm{train}}$ to obtain a HR target image $\overline{\mathbf{x}}$. The "SR model" represents the SR method to be evaluated.

factor z to integer zoom levels (e.g., $z = 19/20$), which we call *pseudo-real* images. Thereby, we effectively create a zoom pyramid [17] of reference images, which we evaluate and report on. Since STMs and STEMs capture images point by point, we select nearest-neighbor interpolation [17] as realistic degradation function for downscaling, which uses no filters and therefore preserves pixel-based noise. Using this concept, we deploy the following two evaluation setups, shown in Fig. 4.

Standard evaluation The standard evaluation (top part in Fig. 4) follows standard supervised SR works [9, 11, 22–24, 27, 32, 39, 40], in evaluating on degraded versions of the HR reference images, although using a noise-preserving degradation. Since test images are paired, the reconstruction of slight irregularities in the periodic structure of the measurements is reflected in the evaluation. Standard evaluation takes HR raw images from a dataset $\mathcal{D}^{\mathrm{test}}_{\mathrm{sL}^{\mathrm{train}}}$ with scale parameter $sL^{\mathrm{train}} = L$, meaning that evaluation is performed solely based on physically captured images with an s times higher scale than what the training dataset $\mathcal{D}^{\mathrm{train}}$ is based on. HR targets $\overline{\mathbf{x}}$ are cropped from the raw image $\breve{\mathbf{x}}$ and from pseudo-real downscaled versions thereof ($\tilde{\mathbf{x}}$). The HR targets $\overline{\mathbf{x}}$ are then degraded by a super-resolution factor of s to LR test inputs $\mathbf{x}$, using once again nearest neighbor interpolation, which results in an approximation of captured images of an s times lower scale. Ultimately, quality metrics for image similarity of target image $\overline{\mathbf{x}}$ and the model's prediction $\mathbf{y}$ are calculated.

Physical evaluation The physical evaluation (bottom part in Fig. 4) can also be carried out for periodic images and is based on *two physically captured unpaired datasets*, a low-resolution one for SR model inputs and a high-resolution one for reference. Thereby, it incorporates real-world microscope image degradation in the evaluation but also requires an alignment of a super-resolved image crop $\mathbf{y}$ to a (pseudo-)real image $\breve{\mathbf{x}}$ ($\tilde{\mathbf{x}}$) of the same zoom level. To accomplish this, test images from a dataset $\mathcal{D}^{\mathrm{test}}_{\mathrm{L}^{\mathrm{train}}}$ with scale parameter $L^{\mathrm{train}} = L$ are the basis for SR model inputs $\mathbf{x}$ (bottom part), while target images $\overline{\mathbf{x}}$ are obtained from a dataset $\mathcal{D}^{\mathrm{test}}_{\mathrm{sL}^{\mathrm{train}}}$ of an s times higher scale parameter (from the top part). Correspondingly to the pseudo-real ground truth images, pseudo-real variants of $\mathcal{D}^{\mathrm{test}}_{\mathrm{L}^{\mathrm{train}}}$ are generated by nearest-neighbor interpolation with factor z, from which SR model input image $\mathbf{x}$ can be cropped. To obtain the target image $\overline{\mathbf{x}}$, the model's output $\mathbf{y}$ is matched against a raw (or pseudo-real) image $\breve{\mathbf{x}}$ ($\tilde{\mathbf{x}}$) taken from $\mathcal{D}^{\mathrm{test}}_{\mathrm{sL}^{\mathrm{train}}}$ to find a matching area used as ground truth $\overline{\mathbf{x}}$. For that, we developed a noise-robust alignment and cropping method, which slightly corrects the image $\breve{\mathbf{x}}$ ($\tilde{\mathbf{x}}$) regarding pixel size and orientation, since the atomic lattices on two point-scanning microscope images are usually not perfectly aligned to each other, as well as actual magnification may slightly differ. A matching crop $\overline{\mathbf{x}}$ is selected via maximizing structural similarity (SSIM) [35] from crop candidates, which are proposed by maximizing Pearson correlation computed for all possible areas based on smoothed versions of the images. Finally, the quality metrics for image similarity of the identified target image $\overline{\mathbf{x}}$ and the model's prediction $\mathbf{y}$ are calculated. Consequently, the physical evaluation assesses the SR results between two unpaired physically recorded datasets, that differ in scale by s.

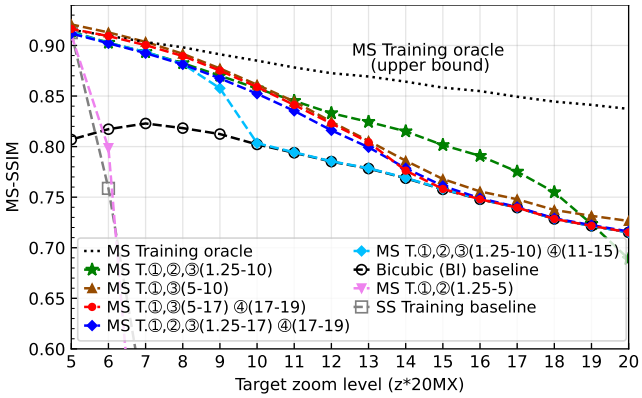


Figure 5. **STEM MS-SSIM** results with **standard evaluation** of a **4x SR** with our proposed multi-scale (MS) training vs. a single-scale (SS) and a bicubic interpolation (BI) baseline, depending on the target zoom level with millionfold magnification (MX) of microscope images of **gallium nitride** (GaN). MS and SS models, the oracle excluded, were trained based on 5MX STEM images $\mathcal{D}^{\text{train}}_{\text{STEM}-5\text{MX}}$.

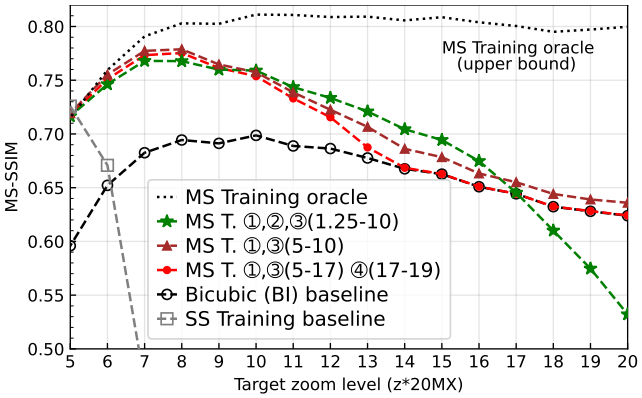


Figure 6. **STEM MS-SSIM** results with **physical evaluation** of a **4x SR** with our proposed multi-scale (MS) training vs. a single-scale (SS) and a bicubic interpolation (BI) baseline, depending on the target zoom level with millionfold magnification (MX) of microscope images of **gallium nitride** (GaN). MS and SS models, the oracle excluded, were trained based on 5MX STEM images $\mathcal{D}^{\text{train}}_{\text{STEM}-5\text{MX}}$.

5. Evaluation and Discussion

Models We conduct experiments on 2-fold super-resolution (2xSR) and on 4-fold super-resolution (4xSR) on the datasets presented in Section 4, Table 2. For all experiments, the lightweight `SwinIR` [23] model topology with 0.878M parameters is used. Details for MS-trained models are given in Table 1. A single-scale (SS) trained model, solely using training data of scale L^{train} as training targets, serves as a baseline (i.e., sub-process ① only). We choose bicubic interpolation (BI) [4] as a non-trainable baseline method, since it is the established method for comparison in single-image SR [12, 25]. For further context, we also trained models on *HR data* using our multi-scale (MS) training. Those MS trainings are based on 20MX STEM images (and nearest neighbor-downscaled to 5MX...19MX) and 12nm STM images (and nearest neighbor-downscaled to 13nm...24nm). As this data is not available in our task definition, we label results with "MS Training oracle".

Training details For model training, we follow Liang et al. [23], employing the AdamW optimizer for training 150k iterations using a stepwise learning rate scheduler with an initial learning rate of 0.001. We use the L1 pixel loss and a batch size of 16 for 4xSR and of 4 for 2xSR. We train on target image crops of 256x256 pixels. The models are trained with `PyTorch` [28] on an `NVidia GTX 1080 Ti` GPU. For the up- and downscaling, we use the implementations of the `Python Pillow Library` [8].

5.1. Evaluation for STEM Microscopy

For STEM data, we evaluate on 4x SR over a range of target zoom levels spanning 5MX to 20MX, based on physically acquired HR 20MX STEM images $\mathcal{D}^{\text{test}}_{\text{STEM}-20\text{MX}}$ including a range of pseudo-real HR downsampled versions thereof (5MX...19MX). Accordingly, the SR model input zoom levels are in the range 1.25MX to 5MX, being created by 4x degradation of the reference images for standard evaluation, or physically captured at 5MX (and downscaled) for physical evaluation. Within this setting, we show the effects of sub-processes in Table 3 and Fig. 5. The SS training baseline (□) uses solely training targets from 5MX images and it's performance for zoom levels above 5MX collapses. By including sub-processes ① and ② ("MS T.①,②(1.25-5)", ▼), the method uses training targets from the original and downscaled images in the fashion of [2, 30], but almost no improvement is apparent. In contrast, using sub-processes ① and ③ ("MS T.①,③(5-10)", ▲), thereby using training targets from original and upscaled images, leads to strong improvements for the configured zoom levels (5MX...10MX) and even generalizes beyond those with small improvements over BI performance.

Combining down- and upscaling for training target generation ("MS T.①,②,③(1.25-10)", ★) improves the performance even further for zoom levels of 10MX to 18MX, while the maximum upscaling zoom level used in training was 10MX. We select this as our main configuration. Interestingly, here, sub-process ② improves results, while it has no noticeable effect in isolation. Unfortunately, it also causes the performance to drop below the BI baseline for zoom ranges above 18MX (★). To counteract this, sub-process ④(11-15) can be included into the configuration (◆), so the model converges to the BI baseline at 10MX, showing the strong impact of sub-process ④. Using sub-process ④(17-19), but now with a higher upscaling maximum in sub-processes ①,②,③(1.25-17) (◆), the performance gets better, approaching "MS T.①,③(5-17) ④(17-

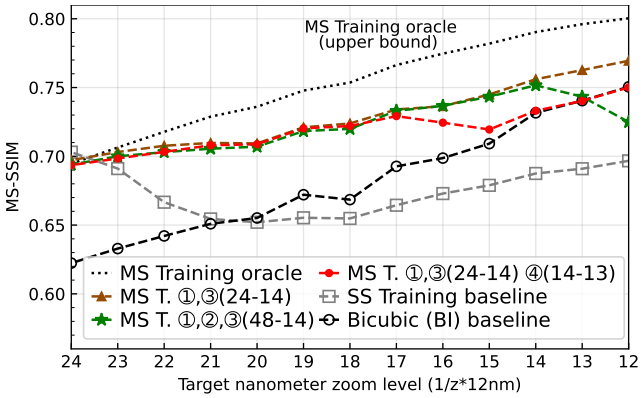


Figure 7. **STM MS-SSIM** results with **standard evaluation** of a **2x SR** with our proposed multi-scale (MS) training vs. a single-scale (SS) and a bicubic interpolation (BI) baseline, depending on the target nanometer zoom level (nm) of microscope images of **graphite**. MS and SS models, the oracle excluded, were trained based on 24nm STM images $\mathcal{D}^{\text{train}}_{\text{STM}-24\text{nm}}$.

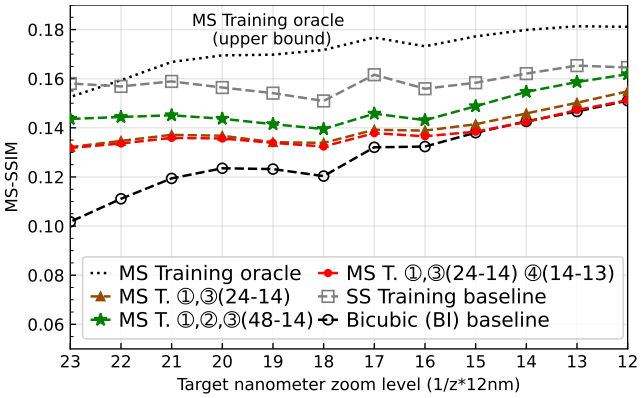


Figure 8. **STM MS-SSIM** results with **physical evaluation** of a **2x SR** with our proposed multi-scale (MS) training vs. a single-scale (SS) and a bicubic interpolation (BI) baseline, depending on the target nanometer zoom level (nm) of microscope images of **graphite**. MS and SS models, the oracle excluded, were trained based on 24nm STM images $\mathcal{D}^{\text{train}}_{\text{STM}-24\text{nm}}$.

Table 3. **Ablations on sub-processes contributions**. PSNR in dB and MS-SSIM for **4xSR** on 20MX STEM (**GaN**) images (400%) and pseudo-real images thereof (200% and 300%) with nearest-neighbor degradation (standard evaluation). The SS baseline ① is trained on single-scale 5MX data, while multi-scale (MS) trainings add the sub-processes: ②(1.25-5); ③(5-10); ④(11-15).

Method	Sub-processes ① ② ③ ④	PSNR / MS-SSIM at % magnification 200%	300%	400%	Tag
Bicubic		21.97 / 0.8021	22.13 / 0.7577	22.15 / 0.7151	○
SS Baseline	✓	19.67 / 0.4272	19.56 / 0.4094	19.96 / 0.4770	□
+ MS	✓ ✓	17.90 / 0.1037	18.52 / 0.2343	19.18 / 0.3674	▼
+ MS	✓ ✓	23.01 / **0.8612**	22.39 / 0.7679	**22.47** / **0.7265**	▲
+ MS	✓ ✓ ✓	**23.03** / 0.8575	**22.86** / **0.8015**	21.58 / 0.6889	★
+ MS	✓ ✓ ✓ ✓	22.96 / 0.8541	22.15 / 0.7586	22.17 / 0.7160	◆

19)" (●). As interesting insight, we note that models using ④ even mimic BI performance for zoom levels above those configured in training, indicating that a generalization of the BI algorithm was learned for unknown zoom level ranges (cf. Fig. 5, ◆, ◆, ●), ensuring that image quality does not fall below the bicubic baseline (cf. Fig. 9, Bicubic vs. MS Train.+ Conv. at 20MX). Fig. 5 concludes that the proposed MS method is able to surpass (or match) the baselines at all target zoom levels.

Fig. 6 confirms the results for *physically captured* HR ground truth images. The same applies to PSNR results in Fig. 1 (and Table 3), where the proposed multi-scale training method (★) is able to surpass the bicubic baseline up to 18MX, or up to 14MX (●) while beyond converging to the BI baseline. *We infer that multi-scale training is beneficial for resolving STEM images up to 360% (18MX/5MX) of highest available zoom level (5MX)* (★), taking advantage of image upscaling for training target generation. In practice, this implies capturing an image at 4.5MX and super-resolve it to 18MX with the 4x SR model.

5.2. Evaluation for STM Microscopy

As a simple STM was used for capturing our STM data, the images of graphite surface structure contain more microscope noise (cf. Fig. 3), leading to an overall more challenging dataset. Therefore, we evaluate the 2x SR task on HR 12nm STM reference images $\mathcal{D}^{\text{test}}_{\text{STM}-12\text{nm}}$ including a range of pseudo-real downscaled versions thereof (24nm...13nm). For STMs, a lower nanometer zoom level corresponds to a higher magnification. In Fig. 7, the proposed MS-trained methods (★ and ●) are able to surpass (or match) the BI baseline at all target nanometer zoom levels, as well as the SS baseline, except at 24nm target nanometer zoom level. Interestingly, for this noisy dataset, sub-process ② does not lead to better model results. Since the graphite STM images contain such high noise level, we suspect the SR of the atomic structure to get more difficult with downscaling of the HR reference images, leading to an overall increasing trend in MS-SSIM for lower nanometer zoom levels, where less downscaling is involved.

Physical evaluation in Fig. 8 shows very low MS-SSIM results, which could be explained by the huge difference in image quality and microscope noise between the two physically captured datasets at 24nm $\mathcal{D}_{\text{STM}-24\text{nm}}$ and 12nm $\mathcal{D}_{\text{STM}-12\text{nm}}$ (cf. Fig. 10, HR ground truth and LR input), making it difficult to evaluate the real-world degradation. Fig. 10 shows a qualitative comparison of an example 2x super-resolved image crop at 18nm for the MS training ①,②,③(48-14)(★), MS training + convergence (●), and the BI and SS baselines. While the SS baseline reaches metric scores above our methods, the qualitative results reveal that the SS model produces very pixelated structures and draws pixel values towards the average gray value (see Fig. 10, histograms). In the context of a noisy reference image,

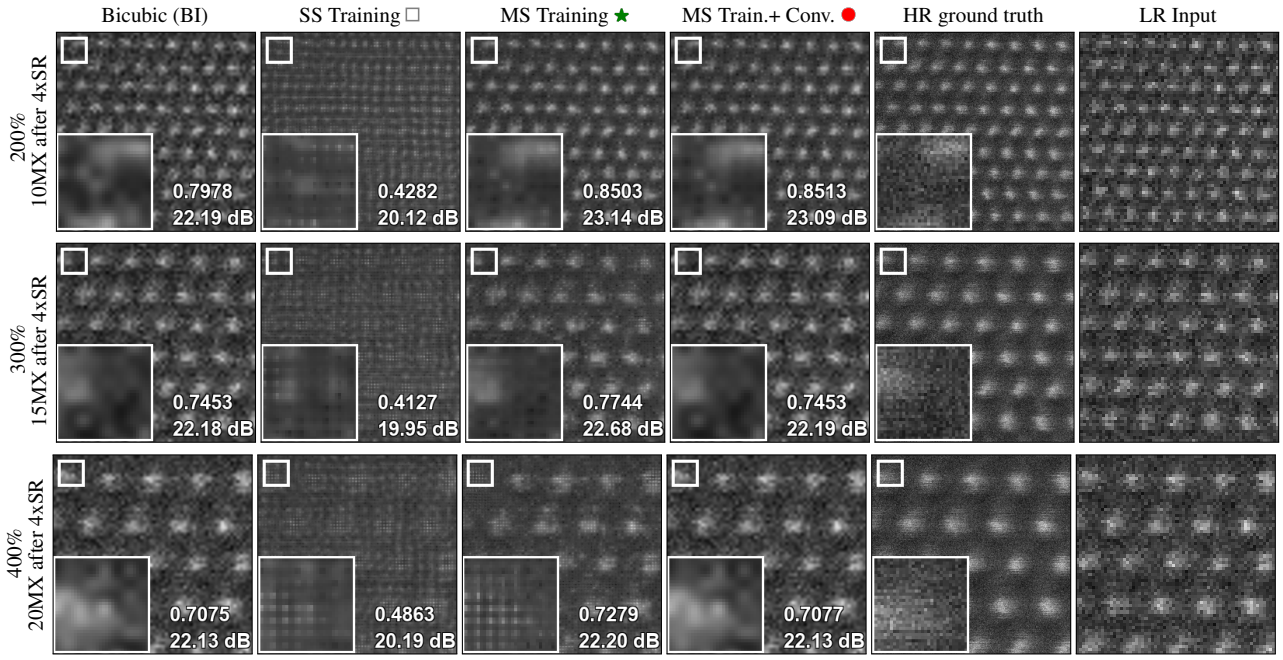


Figure 9. **Qualitative comparison** on STEM (GaN) samples with **4xSR** and **standard evaluation**. Rows display the results for target zoom levels of 10MX (top), 15MX (middle) and 20MX (bottom), being a 200%, 300% and 400% magnification compared to the 5MX (100%) training data $\mathcal{D}^{\text{train}}_{\text{STEM}-5\text{MX}}$. Columns display method results as well as the HR ground-truth crops and the 4x degraded LR input images. HR ground-truth image is taken from the 20MX dataset $\mathcal{D}^{\text{test}}_{\text{STEM}-20\text{MX}}$. At 10MX and 15MX, the MS training results appear closest to the HR ground truth. The MS training + convergence produces visually very similar results to BI for 15MX and 20MX, as configured. MS-SSIM and PSNR results for the methods' output vs. the ground truth are displayed in the right bottom corner (here: metrics of example crop; for test dataset average see Figs. 1, 5). The lower left rectangle is a zoom of the small marked area. Best viewed digitally with zoom.

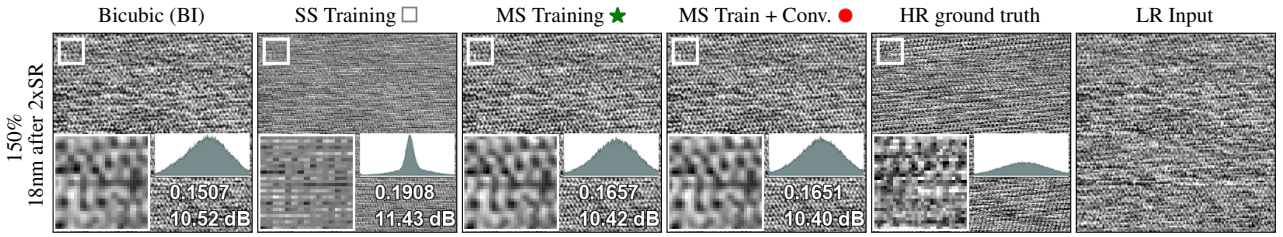


Figure 10. **Qualitative comparison** on STM (graphite) samples with **2xSR** and **physical evaluation**. Columns display method results as well as the HR ground-truth crop and the LR input crop. The LR input image crop from the 24nm dataset $\mathcal{D}^{\text{test}}_{\text{STM}-24\text{nm}}$ is downscaled to 36nm, then super-resolved to 18nm and individually matched to a pseudo-real HR ground truth image obtained from $\mathcal{D}^{\text{test}}_{\text{STM}-12\text{nm}}$ by downscaling to 18nm. All results match individually to the same ground truth. We show image histograms, MS-SSIM and PSNR results in the right bottom corner. The lower left rectangle is a zoom of the small marked area. Best viewed digitally with zoom.

average pixel values can push quantitative metrics, while actually weakening image quality (cf. Fig. 10, SS Training). Due to the overall low MS-SSIM level, this effect might lift the SS baseline above our models in Fig. 8. In contrast, our method (★) preserves the intensities and produces images with smoother structure. Visually comparing results, MS training leads to better quality images than the SS baseline, arguably even better than the physically captured HR ground truth (Fig. 10). Integrating ④ for BI convergence also works for this data, comparing BI and MS Train.+ Conv. (●) in Fig. 8 and image qualities in Fig. 10.

6. Conclusions

In this work, we enable image SR for scanning microscope images beyond the zoom level available for training. Our proposed multi-scale training method is solely based on augmentation of the available LR training data and leverages image upscaling for LR/HR training pair generation. We incorporate real microscope degradation in the evaluation and show improved image quality for zoom levels not available for training. We claim this setup to be a proof of concept, increasing image resolution to surpass the technical limits of STEM and STM microscopy.

References

[1] Waqar Ahmad, Hazrat Ali, Zubair Shah, and Shoaib Azmat. A New Generative Adversarial Network for Medical Images Super Resolution. *Scientific Reports*, 12:9533, June 2022. 1

[2] Namhyuk Ahn, Jaejun Yoo, and Kyung-Ah Sohn. SimUSR: A Simple But Strong Baseline for Unsupervised Image Super-Resolution. In *Proc. of CVPR Workshops*, pages 1953–1961, Seattle, WA, USA, June 2020. 2, 6

[3] Gerd Binnig, Heinrich Rohrer, Christoph Gerber, and Edmund Weibel. Surface Studies by Scanning Tunneling Microscopy. *Physical Review Letters*, 49(1):57–61, July 1982. 1

[4] Burge Burger. *Principles of Digital Image Processing. Core Algorithms*. Springer, Apr. 2009. 3, 6

[5] Chang Chen, Zhiwei Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera Lens Super-Resolution. In *Proc. of CVPR*, pages 1652–1660, Long Beach, CA, USA, June 2019. IEEE. 2

[6] Honggang Chen, Xiaohai He, Linbo Qing, Yuanyuan Wu, Chao Ren, and Ce Zhu. Real-World Single Image Super-Resolution: A Brief Review. *Information Fusion*, 79:124–145, Mar. 2022. 2

[7] Xiangyu Chen, Xintao Wang, Jiantao Zhou, and Chao Dong. Activating More Pixels in Image Super-Resolution Transformer. arXiv:2205.04437 preprint, May 2022. 2

[8] Alex Clark. Pillow (pil fork) documentation. 2015. 6

[9] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-Order Attention Network for Single Image Super-Resolution. In *Proc. of CVPR*, pages 11057–11066, Long Beach, CA, USA, June 2019. IEEE. 2, 3, 5

[10] Kevin de Haan, Zachary S. Ballard, Yair Rivenson, Yichen Wu, and Aydogan Ozcan. Resolution Enhancement in Scanning Electron Microscopy Using Deep Learning. *Scientific Reports*, 9:12050, Aug. 2019. 1, 2, 3

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a Deep Convolutional Network for Image Super-Resolution. In *Proc. of ECCV*, pages 184–199, Zurich, Switzerland, Sept. 2014. 2, 3, 5

[12] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 38(2):295–307, July 2016. 2, 6

[13] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *Proc. of ICLR*, pages 1–21, May 2021. 2

[14] Claude E. Duchon. Lanczos Filtering in One and Two Dimensions. *Journal of Applied Meteorology and Climatology*, 18(8):1016–1022, Aug. 1979. 3

[15] Jeffrey M. Ede and Richard Beanland. Partial Scanning Transmission Electron Microscopy with Deep Learning. *Scientific Reports*, 10(1):8332, May 2020. 2, 3

[16] Linjing Fang, Fred Monroe, Sammy Weiser Novak, Lyndsey Kirk, Cara R. Schiavon, Seungyoon B. Yu, Tong Zhang, Melissa Wu, Kyle Kastner, Alaa Abdel Latif, Zijun Lin, Andrew Shaw, Yoshiyuki Kubota, John Mendenhall, Zhao Zhang, Gulcin Pekkurnaz, Kristen Harris, Jeremy Howard, and Uri Manor. Deep Learning-Based Point-Scanning Super-Resolution Imaging. *Nature Methods*, 18(4):406–416, Apr. 2021. 1, 3

[17] Rafael C. Gonzales and Richard E. Woods. *Digital Image Processing*. Prentice Hall, 2008. 3, 5

[18] Richard Wesley Hamming. *Digital Filters 2nd ed.* Prentice-Hall, 1983. 3

[19] Kyle P. Kelley, Maxim Ziatdinov, Liam Collins, Michael A. Susner, Rama K. Vasudevan, Nina Balke, Sergei V. Kalinin, and Stephen Jesse. Fast Scanning Probe Microscopy via Machine Learning: Non-Rectangular Scans with Compressed Sensing and Gaussian Process Optimization. *Small*, 16(37):2002878, 2020. 2

[20] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-Recursive Convolutional Network for Image Super-Resolution. In *Proc. of CVPR*, pages 1637–1645, June 2016. 2

[21] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic Single Image Super-Resolution Using a Generative Adversarial Network. In *Proc. of CVPR*, pages 4681–4690, Honolulu, HI, USA, July 2017. 2

[22] Jaewon Lee and Kyong Hwan Jin. Local Texture Estimator for Implicit Representation Function. In *Proc. of CVPR*, New Orleans, Lousiana, USA, June 2022. 2, 3, 5

[23] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image Restoration Using Swin Transformer. In *Proc. of ICCV*, pages 1833–1844, virtual, Oct. 2021. 2, 3, 5, 6

[24] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced Deep Residual Networks for Single Image Super-Resolution. In *Proc. of CVPR Workshops*, July 2017. 2, 3, 5

[25] Tairan Liu, Kevin de Haan, Yair Rivenson, Zhensong Wei, Xin Zeng, Yibo Zhang, and Aydogan Ozcan. Deep Learning-Based Super-Resolution in Coherent Imaging Systems. *Scientific Reports*, 9(1):3926, Mar. 2019. 3, 6

[26] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows. In *Proc. of ICCV*, pages 10012–10022, virtual, Oct. 2021. 2

[27] Yiqun Mei, Yuchen Fan, and Yuqian Zhou. Image Super-Resolution With Non-Local Sparse Attention. In *Proc. of CVPR*, virtual, June 2021. 2, 3, 5

[28] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Proc. of NeurIPS*, pages 8024–8035, Vancouver, BC, Canada, Dec. 2019. 6

[29] David Salomon. *Data Compression: The Complete Reference*. Springer Science & Business Media, 2004. 4

[30] Assaf Shocher, Nadav Cohen, and Michal Irani. Zero-Shot Super-Resolution Using Deep Internal Learning. In *Proc. of*

CVPR, pages 3118–3126, Salt Lake City, UT, June 2018. 2, 6

[31] Ying Tai, Jian Yang, and Xiaoming Liu. Image Super-Resolution via Deep Recursive Residual Network. In *Proc. of CVPR*, pages 2790–2798, Honulu, HI, USA, July 2017. 2

[32] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image Super-Resolution Using Dense Skip Connections. In *Proc. of CVPR*, pages 4799–4807, Honulu, HI, USA, July 2017. 2, 3, 5

[33] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is All You Need. In *Proc. of NIPS*, pages 1–11, Long Beach, CA, USA, Dec. 2017. 2

[34] Fan Wang, Dong Yin, and Ruiyuan Song. Image Super-Resolution Using Only Low-Resolution Images. *The Visual Computer*, Aug. 2022. 2

[35] Zhou Wang, Alan Conrad Bovik, Hamid Rahim Sheikh, and Eero P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Trans. on Image Processing*, 13(4):600–612, Apr. 2004. 5

[36] Zhou Wang, Eero Simoncelli, and Alan Bovik. Multi-Scale Structural Similarity for Image Quality Assessment. In *Proc. of ACSSC*, pages 1398–1402, Pacific Grove, CA, USA, Nov. 2003. 4

[37] Roland Wiesendanger. *Scanning Probe Microscopy and Spectroscopy*. Cambridge University Press, Cambridge, Sept. 1994. 1

[38] David B. Williams and C. Barry Carter. *Transmission Electron Microscopy*. Springer, 2009. 1

[39] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss, editors, *Proc. of ECCV*, pages 294–310, Munich, Germany, Sept. 2018. 2, 3, 5

[40] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual Dense Network for Image Super-Resolution. In *Proc. of CVPR*, pages 2472–2481, Salt Lake City, UT, USA, June 2018. 2, 3, 5