Low-Resolution-Only Microscopy Super-Resolution Models Generalizing to Non-Periodicities at Atomic Scale

Björn Möller[1] Zhengyang Li [1] Markus Etzkorn [2] Tim Fingscheidt[1]
{bjoern.moeller, zhengyang.li, m.etzkorn, t.fingscheidt}@tu-bs.de
Technische Universität Braunschweig, Germany [1]Institute for Communications Technology [2]Institute of Applied Physics

Abstract

Super-resolution (SR) methods can accelerate microscopy image capturing and improve quality. Yet, training data is often scarce and low in variability, leading to overfitting models that fail to preserve unseen image structures. Therefore, in this work, we investigate SR model generalization in a low-resource domain, here: material science. First, we propose a training pipeline based on PixMix augmentation for microscopy SR using low-resolution only training data to generate pseudo LR/HR training pairs. The augmentation introduces variability into training images by blending them with high-detail out-of-domain images. Second, using scanning transmission electron microscopy (STEM) images, we show that our proposed training pipeline improves the SR model generalization for `non-periodic high-resolution` *test data of crystalline atomic structures, even if only* `periodic low-resolution` *data is used for training. Furthermore, our proposed pipeline enables STEM SR models to generalize to images with noise characteristics from an unseen recording session. Third, we investigate effects of mixing augmentation strength. Finally, we validate the usage of PixMix on a more comprehensive STEM dataset. Our results demonstrate that frequent image mixing utilizing high-detail out-of-domain data improves SR generalization within low-resource domains such as atomic-scale STEM images of non-periodic matter. Data and code is available*[1].

1. Introduction

Microscopy image super-resolution based on deep learning methods is increasingly gaining traction in natural sciences, such as physics [13, 16], biology [5, 14, 22, 25] and medicine [1, 4] applications. Applied to electron microscopes, using partial scans [7, 12, 13] can reduce the acquisition time as well as the damage to the sample caused by the electron beam. However, most SR models learn to reconstruct a high-resolution (HR) target image from a low-resolution (LR) input image, and thereby depend on high-quality HR images, which cannot always be acquired. Few approaches address the problem of unavailable HR electron microscopy images [14, 20]. For instance, image properties similar to those of a higher-resolution microscope can be leveraged [14], but microscope substitution is not always an option. Another approach [20] trains a SR model with pseudo LR/HR training pairs generated from the available STEM [30] data, seeking to super-resolve beyond highest available training resolution. Yet, it is essential that the obtained SR model generalizes to unseen sample structures and unseen noise characteristics, which has not been shown, and can drastically fail (see Fig. 1b). We adopt this LR-only

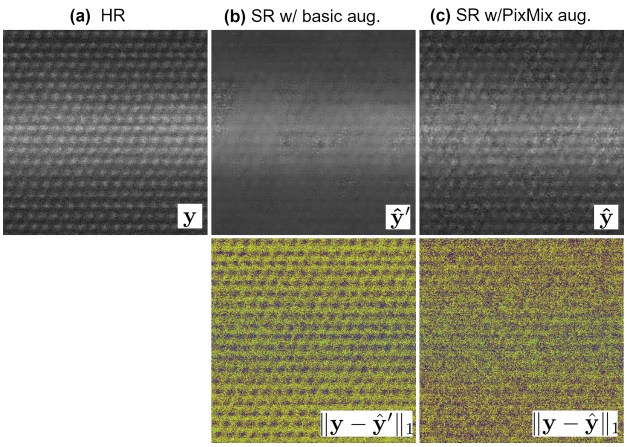


Figure 1. **4x SR model results for atomic scale STEM images: (a)** HR reference, **(b)** SR result with basic augmentations, **(c)** SR result with PixMix augmentations ($K^{\max} = 30, \beta = 2$) during training. Models were trained on periodic 20pm images (low resolution (LR)) and tested on non-periodic 5pm images (high resolution (HR) output). Shown is a challenging case, where **(b)** largely fails the image reconstruction due to non-periodicities and due to noise characteristics from an unseen recording session. The bottom row shows the pixel-wise absolute difference between HR ground-truth and SR result.

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

training strategy and consider these aspects in our evaluation on HR atomic-scale STEM images.

Capturing high-quality STEM images at atomic scale is challenging and costly, often resulting in scarcity of training data. Also, data is often low in variability due to highly repetitive patterns in the examined samples, both making the microscopy SR model prone to overfitting. Image augmentations are used to mitigate this, so image rotation and flipping [26] became common practice. Yet, these image transformations do not introduce new frequency information and we found that they are not sufficient to generalize SR models to unseen microscope noise characteristics (see Fig. 1b). In domains prone to sensor noise, noise injection [14] can be utilize to increase training variability and combat overfitting, however, only augmenting the LR training input. In our work, we randomly augment the training target from which the training input is generated. We introduce variability by image transformations and image mixing, impacting the spatial and frequency domain. For this, we use PixMix [18], an iterative process of mixing the training image with a randomly transformed version of itself or with a mixing image from a high-detail out-of-domain dataset. The method introduces new complexity into the training and can easily be configured in mixing frequency and intensity, which we leverage to study the generalizing effects of augmentation strength to non-periodicities and recording session-specific noise characteristics.

In our experiments, we train on images showing the atomic lattice of a gallium nitride (GaN) crystal. As crystalline materials have a periodic arrangement of atoms in a lattice expanding in all three spatial directions, the corresponding nano-scale images show little variability. Only during evaluation, we introduce images showing non-periodic regions, resulting from a GaN/InGaN heterostructure in the crystal's z-axis, visible as regions of higher brightness (see Fig. 1a). Furthermore, we separate the test images according to their recording sessions to evaluate generalization to noise characteristic from an unseen recording session. We use noise characteristics as a summarizing term for effects on image quality such as microscopy system settings, sample quality, and noise patterns.

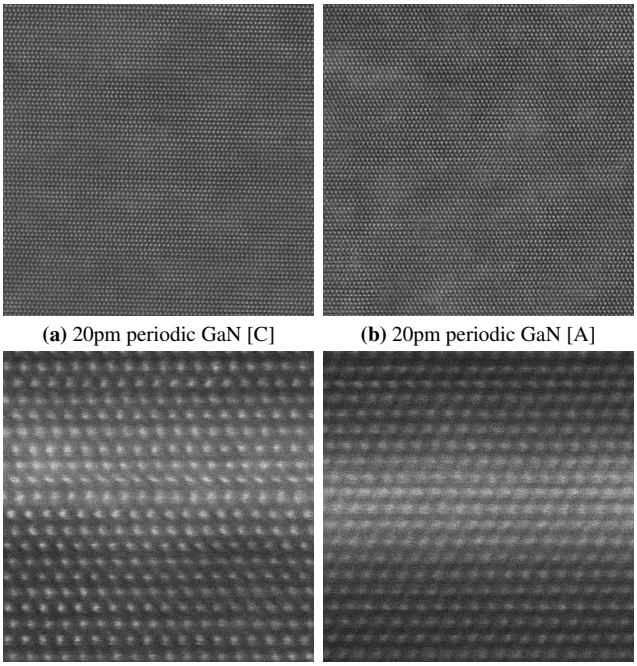
In this paper, we first propose a SR training pipeline based on PixMix [18] augmentation in a LR-only training setting, being the first to explore PixMix for SR and also in the field of microscopy. Second, we train on periodic LR STEM images and evaluate on non-periodic HR STEM images, addressing a challenging unsupervised task, in which we show that PixMix improves generalization to non-periodicities and unseen noise characteristics. Third, we investigate the generalizing effects of image mixing frequency and intensity and find that frequent image mixing shows the best generalization trend. Finally, we validate the idea of using PixMix on a more comprehensive publicly available STEM dataset of further materials, magnification levels, and microscope specifics [11].

2. Related Work

Low-resolution-only image SR Only a few approaches consider the concept of training an SR model without HR resolution images [2, 20, 24], implying that the test resolution is higher than the original resolution of the available training data. These approaches generate pseudo LR/HR training pairs by first scaling the original LR image either down [2, 20, 24] or up [20] to obtain the pseudo-HR image, before applying a degradation to generate the pseudo-LR input. Shocher et al. [24] and Ahn et al. [2] only consider macroscopic data. Möller et al. [20] present a multi-scale augmentation approach for training on microscopic data, including STEM images. They evaluate the method only on highly periodic image structures. In practice, however, examining non-periodicities is of particular interest. We adopt this LR-only training concept including multi-scale augmentation [20] and address the generalization of resulting microscopy SR models to non-periodic structures.

Image augmentation for image SR Image augmentation methods [18, 26, 31–33] aim to increase the variability of the training data to improve generalization of deep learning models. Most methods are developed for high-level vision tasks, such as image classification, seeking to increase robustness of a learned representation. In the low-level vision SR task, where local pixel relations are important, basic geometric transformations, such as random cropping, rotation or flipping [26] and color space manipulation were shown to be effective [31]. Methods that strongly disturb spatial information such as CutOut [8], which introduces sharp transitions, harm SR performance. Mixing methods [32, 33], however, show slight improvements [31] and Feng et al. even put Mixup [33] very effectively to use in a limited training data scenario. Hendrycks et al. present PixMix [18] and show good performance on high-level vision tasks, evaluating robustness and safety measures. PixMix combines image transformations and mixing and can be configured in mixing frequency and intensity, which we leverage to explore the effects of augmentation strength during LR-only SR training.

Generalization for microscopy image SR Microscopic measurements often contain more noise relative to macroscopic images, which introduces additional difficulty since a noisy target must be learned from, while at the same time training data is often scarce. For scanning electron microscopy images, Fang et al. [14] augment their LR training inputs using random noise injection as a way to increase variation in SR training. Wang et al. [28] utilize Mixup [33]



(a) 20pm periodic GaN [C] **(b)** 20pm periodic GaN [A]

(c) 5pm np GaN/InGaN [C] **(d)** 5pm np GaN/InGaN [B]

Figure 2. STEM images of atomic crystal lattices: (**a**, **b**) Training samples with a **periodic** GaN structure at 20pm pixel size and (**c**, **d**) test samples with a **non-periodic** (np) GaN/InGaN structure at 5pm pixel size. The recording session ID from which the image originates is shown in brackets.

to increase their dataset size and show improved results for micrographs of butterfly wings. For nano-scale STEM images, Ede [10] deploys Gaussian blurring to weaken high-frequency noise in training targets. In contrast, we introduce new information to our pseudo-HR training target, using PixMix [18].

3. Measurement Domains and Datasets

Brunswick STEM dataset The Brunswick STEM dataset[2] consists of 120 STEM images showing atomic structures of gallium nitride (GaN) and indium gallium nitride (InGaN) at two magnifications. Images were captured using a `JEOL Neoarm F200` electron microscope with aberration correction using a high angle annular dark field (HAADF) detector. Therefore, bright areas correspond to the positions of atomic rows. A benefit of this data is that each image comes with meta information about pixel size and magnification, being ∼20 picometer (20pm) pixel size at 5 millionfold magnification, which we base our SR model training on, and ∼5pm pixel size at 20 millionfold magnification, which we solely used for evaluation (validation and test). A characteristic of this dataset is also that the images were manually divided into subsets that only show

[2]available at https://github.com/ifnspaml/MicroscopyImageSR

Table 1. Relevant data subsets from the **Brunswick STEM dataset**, which is split into STEM images showing solely periodic ("per") atomic lattices of gallium nitride (GaN) and images also showing non-periodic ("np") regions caused by indium gallium nitride (InGaN) layers. Pixel size is given in picometer (pm). Data has been captured in 3 recording sessions A, B, C.

Split properties	Main experiment data		
	train	validation	test
Pixel size	20pm	5pm	5pm
Non-periodicity	✗	✓	✓
Recording session	A,C	C	B,C
# of samples	27	1	10
Notation	$\mathcal{D}^{\text{train}}_{\text{20pm}-\text{per}}$	$\mathcal{D}^{\text{val}}_{\text{5pm}-\text{np}}$	$\mathcal{D}^{\text{test}-\text{B}}_{\text{5pm}-\text{np}}$ $\mathcal{D}^{\text{test}-\text{C}}_{\text{5pm}-\text{np}}$

periodic atomic structures ($\mathcal{D}_{\text{20pm}-\text{per}}$, $\mathcal{D}_{\text{5pm}-\text{per}}$) and those that also show non-periodic structures ($\mathcal{D}_{\text{20pm}-\text{np}}$, $\mathcal{D}_{\text{5pm}-\text{np}}$). Periodic images have been captured on a thin lamella of a GaN crystal viewed along the $[2\bar{1}\bar{1}0]$ zone axis, showing the periodic atomic lattice as a regularly arranged and repeating pattern (see Fig. 2a, b). In some areas the GaN crystal contained approximately 7 atomic rows of InGaN, introducing a non-periodicity within the atomic lattice (brighter region in Fig. 2c, d). Here, non-periodicity refers to the presence of indium atomic layers within the gallium nitride crystal, which is to be distinguished from effects of sample contamination or defects, which are also present in some images, albeit much less pronounced.

All images where captured with 2048x2048 pixels with an exposure time of 3µs/pixel or 6µs/pixel at random but different positions of the sample over three separate recording sessions and contain noise to some degree. For experimentation, we split the data into separate training, validation and test sets. Details for the relevant subsets are shown in Table 1.

Warwick STEM dataset For ablation experiments, we also use the Warwick STEM Crops dataset [11]. It contains 161,069 images of size 512x512 of non-overlapping regions cropped from STEM images. The original data was measured by multiple researchers using a `University of Warwick JEOL ARM 200F` electron microscope between the years 2010 and 2018. It is partitioned in training, validation, and test sets consisting of 110,933, 21,259, and 28,877 images, respectively, each captured by a different subset of scientists. Due to the varying materials and microscope settings, the dataset reveals diverse characteristics. Images are of variable quality, contain noise, and are roughly equally distributed between dark and bright field detectors, while atomic lattices are visible in about two thirds of the images. No meta information about pixel sizes or magnification is provided.

4. Proposed Method (Training Pipeline)

Here, we present our proposed training pipeline, which creates pseudo-LR/HR training pairs and utilizes additional mixing images to introduce variability into the microscopy training data. The pipeline includes PixMix [18], an augmentation method so-far not investigated for SR nor in the microscopy domain. It utilizes a highly randomized process in which potentially transformed versions of the training image and image structures from a mixing dataset are iteratively mixed to a training image. We integrate it in a LR-only training setup as shown in Fig. 3. Input to the pipeline is a grayscale training image $\check{\mathbf{x}} \in \mathbb{G}^{H \times W}$, with $\mathbb{G} = [0, 255]$ being the set of gray values and H, W the height and width, respectively. The method also requires a dataset of mixing images $\mathcal{D}^{\text{mix}}$, for which we use the fractal dataset provided by Hendrycks et al. [18] and convert it to grayscale.

Random multi-scale augmentation and random cropping First, we apply multi-scale augmentation [20] to a training image $\check{\mathbf{x}}$ taken from $\mathcal{D}_{\text{LR}}^{\text{train}}$, scaling it up or down either by bicubic or by nearest-neighbor interpolation respectively. The scaling is done with a randomly drawn factor from a predefined set, leading to image sizes between 25% and 400% of original image size. Followed by random cropping, the precursor image $\overline{\overline{\mathbf{x}}}$ for a pseudo target is created, which is subjected to further PixMix augmentation in our pipeline.

Random bicubic upscaling and random cropping As a prerequisite for PixMix, a mixing image $\check{\mathbf{x}}^{\text{mix}}$ is randomly drawn from $\mathcal{D}^{\text{mix}}$. In order to have more control over the augmentation level, we have included the option to upscale the image with bicubic interpolation. This can be done to reduce detail and soften the sharp edges of the relatively high-detail mixing images to better mimic the diffuse structures typical in STEM images. The upscaling factor b is randomly chosen from a range of integer values, which we set between 5 and 20 in our experiments, yielding in a strong enlargement and blurring of the mixing image. This is followed by random cropping to obtain an image section $\overline{\overline{\mathbf{x}}}^{\text{mix}}$ that corresponds to the size of the training crop $\overline{\overline{\mathbf{x}}}$.

Random addition or multiplication The core idea of PixMix [18] is an iterative mixing of the potentially transformed training image $\overline{\mathbf{x}}_k$ with a transformed version of itself or the mixing crop $\overline{\overline{\mathbf{x}}}_k^{\text{mix}}$ for $K \in \{0, 1, ..., K^{\max}\}$ iterations, while K is randomly chosen and $K = 0$ means the mixing is skipped. Mixing is implemented as addition or multiplication, selected at random. For addition, the images $\overline{\mathbf{x}}_k$ and $\overline{\mathbf{x}}_k^{\text{mix}}$ are weighted by mixing factors m_k and m_k^{mix}:

$$\overline{\mathbf{x}}_{k+1} = m_k \cdot \overline{\mathbf{x}}_k + m_k^{\text{mix}} \cdot \overline{\mathbf{x}}_k^{\text{mix}}. \tag{1}$$

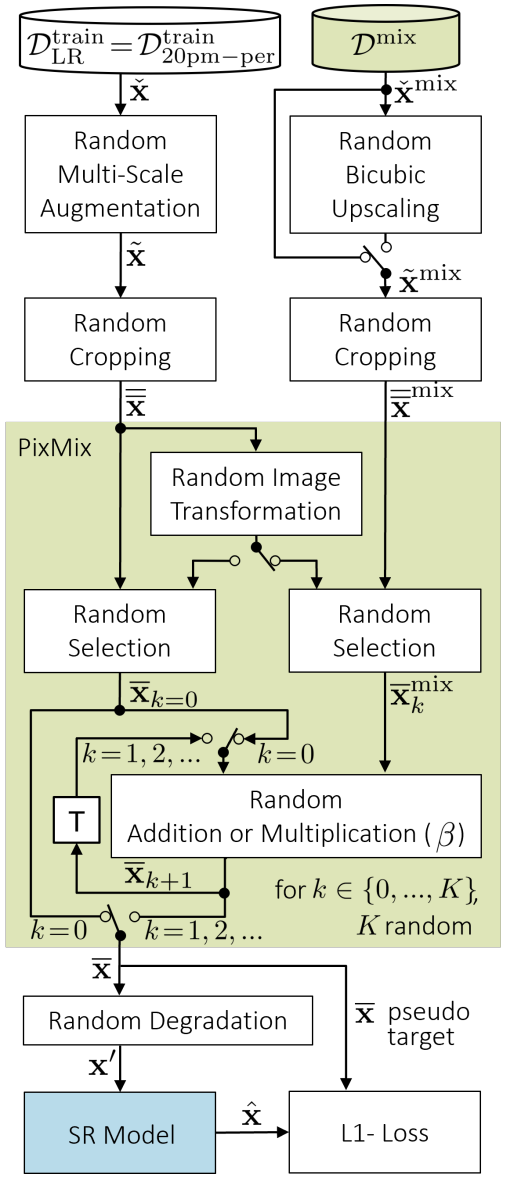


Figure 3. **Proposed training pipeline using PixMix** After multi-scale augmentation and cropping, a training image crop $\overline{\overline{\mathbf{x}}}$ is iteratively mixed with a mixing image crop $\overline{\overline{\mathbf{x}}}^{\text{mix}}$ from the mixing set $\mathcal{D}^{\text{mix}}$. This is done by K iterations of mixing $\overline{\mathbf{x}}_k^{\text{mix}}$ with $\overline{\mathbf{x}}_k$, weighted by factors or exponents independently drawn from beta distributions, while K is also randomly chosen. In each iteration, $\overline{\mathbf{x}}_k^{\text{mix}}$ is generated from $\overline{\overline{\mathbf{x}}}$ or $\overline{\overline{\mathbf{x}}}^{\text{mix}}$, until the pseudo target $\overline{\mathbf{x}}$ is obtained, which is then downscaled by a randomly selected degradation function to create the new pseudo-LR input $\mathbf{x}'$.

For multiplication, the images $\overline{\mathbf{x}}_k$ and $\overline{\mathbf{x}}_k^{\text{mix}}$ are adjusted with m_k and m_k^{mix} as mixing exponents:

$$\overline{\mathbf{x}}_{k+1} = (\overline{\mathbf{x}}_k)^{m_k} \cdot (\overline{\mathbf{x}}_k^{\text{mix}})^{m_k^{\text{mix}}}. \tag{2}$$

For this, m_k and m_k^{mix} are independently drawn from beta distributions, defined by one of the two following randomly

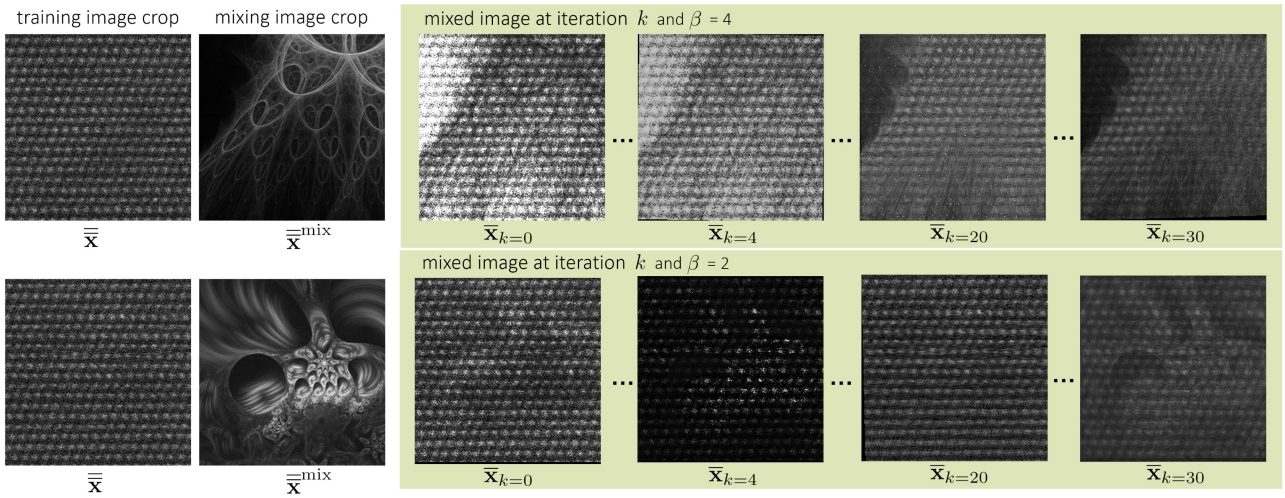


Figure 4. **STEM image transformation and mixing with PixMix**: Two example STEM training image crops $\overline{\overline{\mathbf{x}}}$ (upper and lower panel) that show GaN at 20pm pixel size are repeatedly mixed with one randomly chosen mixing image crop $\overline{\overline{\mathbf{x}}}^{\text{mix}}$ for up to 30 iterations each. The image quality can vary greatly between iterations, especially for lower β values.

selected cases. In the first case, m_k and m_k^{mix} are drawn from complementary beta distributions $m_k \sim \mathcal{B}(\beta, 1)$, $m_k^{\text{mix}} \sim \mathcal{B}(1, \beta)$, implying that a large β decreases the probability for a high mixing impact of $\overline{\overline{\mathbf{x}}}_k^{\text{mix}}$. In the second case, mixing factors are independently drawn from the same distribution ($m_k, m_k^{\text{mix}} \sim \mathcal{B}(1, \beta)$), while m_k is incremented by 1 and the value for m_k^{mix} is negative [18]. In short, β influences the mixing intensity, while K controls mixing frequency.

Random image transformation An image transformation is applied to the training image crop $\overline{\overline{\mathbf{x}}}$ randomly chosen for each iteration from a set of transformation functions. This set consists of contrast normalization, histogram equalization, posterization, rotation, pixel inversion, shearing and translation, most of which are regulated in magnitude by a random chosen value between 10% and 100% [18].

Random degradation To generate a pseudo training input $\mathbf{x}'$, the pseudo target $\overline{\mathbf{x}}$ is degraded with a randomly chosen downscaling function from the set of nearest neighbor [3], Lanczos [9], bilinear [3], bicubic [3], box [15], and Hamming [17] interpolation, also introducing variation into the learned LR/HR mapping [20].

An example of this process is shown in Fig. 4 for two STEM training image crops $\overline{\overline{\mathbf{x}}}$ showing GaN at 20pm pixel size and two randomly chosen mixing image crops $\overline{\overline{\mathbf{x}}}^{\text{mix}}$.

5. Experimental Setup

Training data: periodic and low-resolution We follow the training paradigm of LR-only training, which assumes the absence of good quality HR images to learn from, a scenario naturally occurring in microscopy. Thereby, the available images are interpreted as LR and are the basis for pseudo LR/HR training pair generation to enable supervised SR model training. We also investigate the problem of generalizing to unseen image structures, introduced by non-periodic atomic structures, that do not appear in the training data. Accordingly, we only use $\mathcal{D}_{\text{LR}}^{\text{train}} = \mathcal{D}_{\text{20pm-per}}^{\text{train}}$ for model training.

Training details For model training, the AdamW optimizer is employed for 30k iterations with an initial learning rate of 0.0002 and stepwise reduction, optimizing for L1 pixel loss. The initial learning rate for fine-tuning is reduced to 0.0001. For ablations on the much larger and more comprehensive Warwick STEM dataset, iterations are increased to 150k. We use a batch size of 16 with a pseudo target image size of 256x256 pixels. The models are trained on an `NVidia GTX 2080 Ti` GPU using the `PyTorch` framework [21]. Image augmentations and scaling functions are implemented using the `Python Pillow Library` [6].

Evaluation process We test on the non-periodic 5pm dataset $\mathcal{D}_{\text{5pm-np}}^{\text{test}}$, which was recorded at a four times higher magnification than the training images, corresponding to higher resolution images with an also four times smaller pixel size of 5pm. Test input images are created by downsampling the 5pm images using nearest-neighbor downscaling (see Fig. 5), which is noise preserving and acts as an approximation of the real world degradation between images of two magnifications [20]. For model and parameter selec-

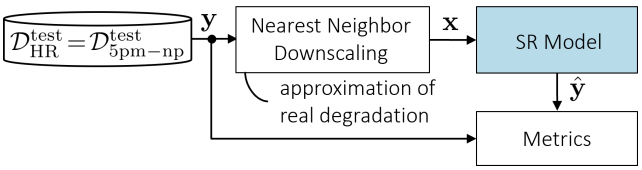


Figure 5. **Evaluation process**: HR test image $\mathbf{y}$ is degraded by nearest-neighbor downscaling to the LR SR model input $\mathbf{x}$. Several image similarity metrics are calculated comparing the original HR image $\mathbf{y}$ and the super-resolved image $\hat{\mathbf{y}}$.

tion, we validate on the corresponding validation split of the non-periodic 5pm dataset $\mathcal{D}_{5\mathrm{pm-np}}^{\mathrm{val}}$.

Metrics For evaluation, we report peak-signal-to-noise ratio (PSNR) [23] and the structural similarity index measure (SSIM) [29] as image similarity metrics. Since STEM images contain high-frequency noise, we also introduce low-pass filtered variants of these metrics. For this, we preprocess the HR images with a 3x3 or 5x5 Gaussian filter to attenuate the effect of these high frequencies and also report the 3x3 and 5x5 HR Gaussian-filtered PSNR and SSIM.

Models and baselines For all experiments, we use the lightweight `SwinIR` [19] model topology with 0.878M parameters, which is composed of transformer blocks [27] using windowed multi-head self-attention. As baseline we train a model using, besides random multi-scale augmentation and cropping, only random horizontal flipping and rotation by 90° as augmentations (basic augmentation). We employ random multi-scale augmentation [20] and cropping for all experiments using the Brunswick STEM dataset. Given a pixel size of 20pm for training, we configure the multi-scale augmentation in the range between 25% and 400%, yielding pixel sizes between 80pm and 5pm. We define a set of 30 scaling values for a random selection, equally distributed between up- and downscaling. Standard PixMix configuration for ImageNet training is $K^{\max}$=4 and β=4 [18], which can be understood as a starting point for experiments on augmentation strength. We also report bicubic interpolation [3] as a non-trainable baseline.

6. Evaluation and Discussion

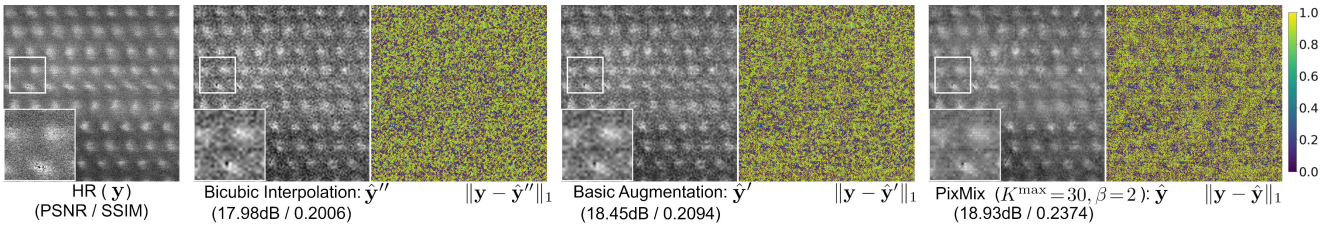
Generalizing to non-periodicities In Table 2 the models trained with periodic 20pm training data are evaluated on the non-periodic 5pm validation set $\mathcal{D}_{5\mathrm{pm-np}}^{\mathrm{val}}$ and the non-periodic 5pm test set $\mathcal{D}_{5\mathrm{pm-np}}^{\mathrm{test}}$. Basic augmentation shows better values for PSNR (19.44 dB vs. 19.07 dB) and SSIM (0.2617 vs. 0.2525) compared to bicubic interpolation. As expected, all methods show higher values for Gaussian-filtered metrics, demonstrating the high noise level in this task. We evaluate on the validation set to find the optimal PixMix [18] hyperparameters $K^{\max}$ and β for our proposed training pipeline. We use a fixed $\beta = 4$ when optimizing for the maximum number of possible mixing iterations $K^{\max} \in \{4, 6, 8, 10, 20, 30, 40\}$. Results show a clear trend that more mixing iterations lead to better quantitative results, yet stagnating at $K^{\max} = 30$, which we continue to use. Our first observation is that increasing $K^{\max}$, which introduces more variability in the spatial and frequency domains, helps to generalize from LR periodic images to HR non-periodic images. Also, upscaling the mixing image by b does not improve results for this data. We then optimize for β, which influences the mixing factors/exponents m_k, m_k^{mix} by shaping their underlying beta distributions. While a lower β increases the impact of the mixing image, $\beta = 1$ marks an extreme case, in which both are drawn from a uniform distribution. For $K^{\max} = 30$, we observe that $\beta = 2$ yields the highest PSNR of 20.05 dB with an improvement of 0.98 dB over bicubic interpolation (19.07 dB). While $\beta = 1$ yields a slightly higher SSIM, PSNR peaks at 20.05 dB with $\beta = 2$. Therefore, we chose $K^{\max} = 30$ and $\beta = 2$ for this dataset. We also show test set performance and find the trend to be confirmed.

Generalizing to recording session-specific noise characteristics In Table 3 we separately test on images from recording sessions C (seen) and B (unseen). Since the training data contains LR images from recording session C, yet captured with a higher pixel size of 20pm, similar noise characteristics could be seen to those of 5pm pixel size test images from $\mathcal{D}_{5\mathrm{pm-np}}^{\mathrm{test-C}}$. In contrast, noise characteristics in images from recording session B ($\mathcal{D}_{5\mathrm{pm-np}}^{\mathrm{test-B}}$) are unseen in training and validation, which allows a separate evaluation of generalization to recording session-specific noise characteristics. Interestingly, PSNR results for bicubic interpolation are ∼2 dB lower for unseen session B (19.46 dB vs. 17.51 dB). In fact, all values for $\mathcal{D}_{5\mathrm{pm-np}}^{\mathrm{test-B}}$ have a significantly lower absolute level, which indicates a more challenging SR task. Basic augmentation surpasses bicubic interpolation for seen session C, however, it fails the image reconstruction for images from unseen session B with a low SSIM of 0.1207, even below bicubic interpolation (0.1553). A comparison of the qualitative results illustrates the problem of missing model generalization (see Fig. 6b, basic augmentation). Here, PixMix improves results and even enables the reconstruction of image structures. Though still imperfect, it demonstrates its potential for generalization (Fig. 6b, PixMix). PixMix ($K^{\max} = 30$, $\beta = 2$) improves PSNR by ∼1 dB (20.42 dB vs. 19.46 dB) for seen session C and by ∼1.4 dB (18.90 dB vs. 17.51 dB) for unseen session B relative to bicubic interpolation. It thereby decreases the mentioned gap in PSNR between recording sessions from ∼2 dB to ∼1.5 dB. We assume, that the SR models overfit to recording session-specific noise characteristics, as noise levels are high and data is scarce, which we mitigate by introducing variability during training.

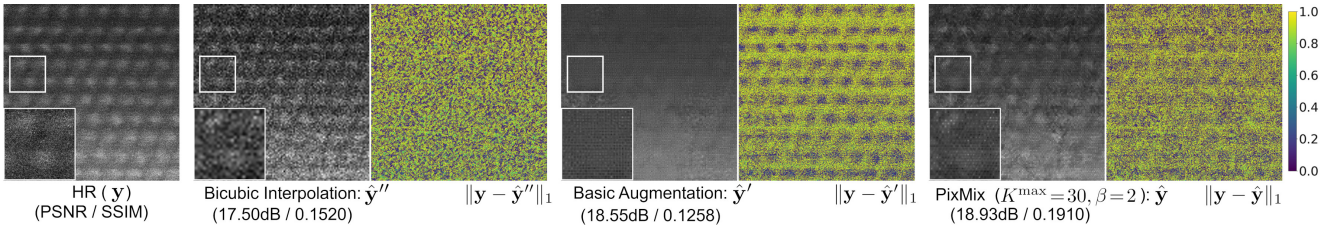
Table 2. **Generalization to non-periodic HR images**: Quantitative results for a 4x SR `SwinIR` lightweight model **trained with various PixMix augmentation strengths**, evaluated on the **5pm non-periodic** $\mathcal{D}^{\text{val}}_{\text{5pm-np}}$ **validation split** of the Brunswick STEM dataset. The 5pm non-periodic **test results** ($\mathcal{D}^{\text{test}}_{\text{5pm-np}}$) are also shown. PSNR (in dB) and SSIM as well as the 3x3 or 5x5 Gaussian-filtered versions are reported. Models were trained from scratch for 30k iterations on periodic 20pm images $\mathcal{D}^{\text{train}}_{\text{20pm-per}}$. Basic augmentation includes random horizontal flipping and rotation by 90°. Parameters $K^{\max}$, β, b control maximum iterations, mixing factors/exponents, and mixing image upscaling in the training pipeline described in Fig. 4. If β is not stated, the mixing crop was not upscaled.

Method	$\mathcal{D}^{\text{val}}_{\text{5pm-np}}$ PSNR	SSIM	$\mathcal{D}^{\text{val-3x3}}_{\text{5pm-np}}$ PSNR	SSIM	$\mathcal{D}^{\text{val-5x5}}_{\text{5pm-np}}$ PSNR	SSIM	$\mathcal{D}^{\text{test}}_{\text{5pm-np}}$ PSNR	SSIM	$\mathcal{D}^{\text{test-3x3}}_{\text{5pm-np}}$ PSNR	SSIM	$\mathcal{D}^{\text{test-5x5}}_{\text{pm-np}}$ PSNR	SSIM
Bicubic interpolation	19.07	0.2525	22.99	0.4485	24.01	0.5020	18.87	0.2331	22.81	0.4229	23.83	0.4750
Basic Augmentation	19.44	0.2617	23.77	0.4834	24.82	0.5416	19.41	0.2284	24.05	0.4681	25.11	0.5448
PixMix [18]												
$K^{\max}$=4, β=4,$b\in\{10,\ldots,20\}$	19.30	0.2647	23.47	0.4749	24.53	0.5301	19.42	0.2522	24.05	0.4777	25.18	0.5429
$K^{\max}$=4, β=4,$b\in\{5,\ldots,10\}$	19.32	0.2656	23.51	0.4771	24.56	0.5323	19.45	0.2528	24.10	0.4797	25.23	0.5452
$K^{\max}$=4, β=4	19.54	0.2739	23.95	0.5000	25.03	0.5582	19.59	0.2591	24.41	0.4968	25.56	0.5634
$K^{\max}$=6, β=4	19.67	0.2783	24.24	0.5140	25.34	0.5744	19.70	0.2624	24.67	0.5084	25.84	0.5775
$K^{\max}$=8, β=4	19.66	0.2771	24.24	0.5130	25.34	0.5735	19.69	0.2613	24.65	0.5074	25.82	0.5758
$K^{\max}$=10, β=4	19.72	0.2794	24.35	0.5189	25.45	0.5801	19.74	0.2632	24.76	0.5129	25.94	0.5831
$K^{\max}$=20, β=4	19.88	0.2851	24.70	0.5359	25.83	0.6002	19.86	0.2673	25.06	0.5272	26.26	0.6005
$K^{\max}$=30, β=4	**20.02**	**0.2889**	**25.07**	**0.5536**	**26.24**	**0.6225**	**19.95**	**0.2699**	**25.31**	**0.5400**	**26.56**	**0.6177**
$K^{\max}$=40, β=4	20.01	0.2888	25.02	0.5521	26.19	0.6206	19.94	0.2696	25.28	0.5389	26.53	0.6168
$K^{\max}$=4, β=3	19.66	0.2775	24.22	0.5124	25.32	0.5725	19.69	0.2622	24.66	0.5085	25.83	0.5776
$K^{\max}$=4, β=2	19.72	0.2796	24.36	0.5190	25.46	0.5801	19.75	0.2634	24.77	0.5129	25.95	0.5828
$K^{\max}$=4, β=1	19.76	0.2809	24.44	0.5240	25.55	0.5860	19.77	0.2645	24.82	0.5182	26.00	0.5894
$K^{\max}$=30, β=6	19.88	0.2850	24.70	0.5356	25.83	0.6001	19.85	0.2669	25.03	0.5258	26.24	0.5997
$K^{\max}$=30, β=4	20.02	0.2889	25.07	0.5536	26.24	0.6225	19.95	**0.2699**	25.31	0.5400	26.56	0.6177
$K^{\max}$=30, β=2	**20.05**	0.2891	**25.16**	0.5578	**26.36**	0.6283	**19.96**	0.2695	**25.35**	**0.5421**	**26.61**	**0.6207**
$K^{\max}$=30, β=1	20.04	**0.2896**	25.13	**0.5595**	26.31	**0.6302**	19.87	0.2689	25.11	0.5391	26.32	0.6171

relative augmentation strength



(a) Non-periodic test image crop from $\mathcal{D}^{\text{test-C}}_{\text{5pm-np}}$ after 4x super-resolution.



(b) Non-periodic test image crop from $\mathcal{D}^{\text{test-B}}_{\text{5pm-np}}$ after 4x super-resolution.

Figure 6. **Qualitative comparison** for a **4x SR** of non-periodic test images from (a) recording session C $\mathcal{D}^{\text{test-C}}_{\text{5pm-np}}$ (seen) and (b) recording session B $\mathcal{D}^{\text{test-B}}_{\text{5pm-np}}$ (unseen). The image crops are taken from regions showing non-periodicities that appear as regions of higher brightness.

Generalization using pre-training A common method aiming to increase generalization is to pre-train a model on a broader dataset and fine-tune that model's weights using the dataset of interest. Following this, we pre-train on the Warwick STEM dataset [11] and fine-tune on the Brunswick STEM dataset. Table 4 reveals that our pipeline using PixMix ($K^{\max}=30$, $\beta=2$) improves PSNR by 0.89 dB compared to bicubic interpolation (19.76 dB vs. 18.87

Table 3. **Generalization to recording noise characteristics**: Quantitative results for a 4x SR `SwinIR` lightweight model **trained with various PixMix augmentation stregths**, tested separately for images from **recording sessions C and B** of the Brunswick STEM dataset **5pm non-periodic test split** ($\mathcal{D}^{\text{test}-\mathbf{C}}_{\text{5pm}-\text{np}}$ and $\mathcal{D}^{\text{test}-\mathbf{B}}_{\text{5pm}-\text{np}}$). PSNR (in dB) and SSIM as well as the 3x3 or 5x5 Gaussian-filtered versions are reported. Models were trained from scratch for 30k iterations on periodic 20pm images $\mathcal{D}^{\text{train}}_{\text{20pm}-\text{per}}$. Basic augmentation includes random horizontal flipping and rotation by 90°. Parameters $K^{\max}$, β, b control iterations, mixing factors/exponents, and mixing image upscaling in the training pipeline described in Fig. 4. If β is not stated, the mixing crop was not upscaled.

	Method	seen session **C**						unseen session **B**					
		$\mathcal{D}^{\text{test}-\mathbf{C}}_{\text{5pm}-\text{np}}$		$\mathcal{D}^{\text{test}-\mathbf{C}-\text{3x3}}_{\text{5pm}-\text{np}}$		$\mathcal{D}^{\text{test}-\mathbf{C}-\text{5x5}}_{\text{5pm}-\text{np}}$		$\mathcal{D}^{\text{test}-\mathbf{B}}_{\text{5pm}-\text{np}}$		$\mathcal{D}^{\text{test}-\mathbf{B}-\text{3x3}}_{\text{5pm}-\text{np}}$		$\mathcal{D}^{\text{test}-\mathbf{B}-\text{5x5}}_{\text{5pm}-\text{np}}$	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
	Bicubic interpolation	19.46	0.2665	23.37	0.4676	24.39	0.5218	17.51	0.1553	21.51	0.3186	22.54	0.3659
relative augmentation strength ↓	Basic Augmentation	19.84	0.2745	24.19	0.5050	25.25	0.5656	18.42	0.1207	23.73	0.3820	24.78	0.4961
	PixMix [18]												
	$K^{\max}$=4, β=4	19.95	0.2883	24.40	0.5220	25.48	0.5818	18.76	0.1910	24.44	0.4380	25.75	0.5204
	$K^{\max}$=6, β=4	20.08	0.2925	24.68	0.5355	25.78	0.5975	18.82	0.1920	24.64	0.4452	25.99	0.5310
	$K^{\max}$=8, β=4	20.07	0.2913	24.68	0.5346	25.79	0.5967	18.79	0.1912	24.56	0.4438	25.90	0.5271
	$K^{\max}$=10, β=4	20.13	0.2938	24.82	0.5417	25.93	0.6050	18.82	0.1918	24.63	0.4457	25.97	0.5320
	$K^{\max}$=20, β=4	20.30	0.2994	25.19	0.5597	26.33	0.6263	18.85	0.1924	24.75	0.4515	26.11	0.5404
	$K^{\max}$=30, β=4	**20.41**	**0.3025**	**25.47**	**0.5740**	**26.64**	**0.6444**	18.90	0.1939	24.94	0.4606	26.35	0.5556
	$K^{\max}$=40, β=4	20.38	0.3021	25.41	0.5719	26.58	0.6421	**18.92**	**0.1940**	**24.98**	**0.4620**	**26.40**	**0.5576**
	$K^{\max}$=30, β=6	20.27	0.2986	25.12	0.5565	26.26	0.6229	18.87	0.1932	24.81	0.4540	26.18	0.5457
	$K^{\max}$=30, β=4	20.41	**0.3025**	25.47	0.5740	26.64	0.6444	**18.90**	0.1939	24.94	0.4606	26.35	0.5556
	$K^{\max}$=30, β=2	**20.42**	0.3020	**25.51**	**0.5764**	**26.70**	**0.6481**	**18.90**	0.1938	**24.97**	0.4621	**26.40**	0.5567
	$K^{\max}$=30, β=1	20.28	0.3008	25.19	0.5712	26.33	0.6423	**18.90**	**0.1944**	24.92	**0.4643**	26.31	**0.5582**

Table 4. **Quantitative results** for a 4x SR `SwinIR` lightweight model, **pre-trained on the Warwick STEM dataset** $\mathcal{D}^{\text{train}}_{\text{warwick}}$ and **fine-tuned** on $\mathcal{D}^{\text{train}}_{\text{20pm}-\text{per}}$. Results for the 5pm non-periodic test split of the Brunswick STEM dataset $\mathcal{D}^{\text{test}}_{\text{5pm}-\text{np}}$ are reported. Fine-tuning was done for 30k iterations.

Method	$\mathcal{D}^{\text{test}}_{\text{5pm}-\text{np}}$		$\mathcal{D}^{\text{test}-\text{3x3}}_{\text{5pm}-\text{np}}$		$\mathcal{D}^{\text{test}-\text{5x5}}_{\text{5pm}-\text{np}}$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic interpolation	18.87	0.2331	22.81	0.4229	23.83	0.4750
Basic Augmentation	19.37	0.2253	24.02	0.4675	25.10	0.5478
PixMix $K^{\max}$=4, β=4	19.38	0.2508	23.97	0.4743	25.10	0.5382
PixMix $K^{\max}$=30, β=2	**19.76**	**0.2667**	**24.77**	**0.5248**	**25.93**	**0.5979**

Table 5. **Quantitative results** for a 4x SR `SwinIR` lightweight model, **trained on the Warwick STEM dataset** $\mathcal{D}^{\text{train}}_{\text{warwick}}$ and evaluated on the Warwick test split $\mathcal{D}^{\text{test}}_{\text{warwick}}$. Models were trained from scratch for 150.000 iterations.

Method	$\mathcal{D}^{\text{test}}_{\text{warwick}}$	
	PSNR	SSIM
Bicubic interpolation	21.86	0.3434
Basic Augmentation	23.33	0.3589
PixMix $K^{\max}$=4, β=4	23.43	0.3925
PixMix $K^{\max}$=30, β=2	**24.64**	**0.3933**

dB). PixMix ($K^{\max} = 30$, $\beta = 2$) also outperforms basic augmentation (19.76 dB vs 19.37dB) confirming the effectiveness of our pipeline when using pre-training.

Application to multifaceted STEM data For further analysis on a more comprehensive and more general microscopy dataset, we conduct experiments on the Warwick STEM dataset [11]. Since it covers various pixel sizes in training and evaluation, it is not a strict LR-only task and we train without multi-scale augmentation. PixMix is also highly effective here, showing a 2.78 dB higher PSNR compared to bicubic interpolation (24.64 dB vs 21.86 dB), and even increases PSNR by 1.31 dB (24.64 dB vs. 23.33 dB) and SSIM by 0.0344 (0.3933 vs 0.3589) over basic augmentation when using $K^{\max} = 30$ and $\beta = 2$.

7. Conclusions

In this work, we improve the generalization of microscopy SR models to non-periodic high-resolution STEM images of atomic structures, given only periodic low-resolution STEM images for training. Our proposed training pipeline utilizes PixMix augmentation to introduce variability into training images by mixing them with high-detail out-of-domain images. We also show improved generalization to unseen noise characteristics. Thereby we introduce PixMix augmentation to SR and microscopy and demonstrate its potential for microscopy SR model generalization.

References

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

[3] Burge Burger. *Principles of Digital Image Processing. Core Algorithms*. Springer, Apr. 2009. 5, 6

[4] Jintai Chen, Haochao Ying, Xuechen Liu, Jingjing Gu, Ruiwei Feng, Tingting Chen, Honghao Gao, and Jian Wu. A Transfer Learning Based Super-Resolution Microscopy for Biopsy Slice Images: The Joint Methods Perspective. *Transactions on Computational Biology and Bioinformatics*, 18(1):103–113, Jan. 2021. 1

[5] Rong Chen, Xiao Tang, Yuxuan Zhao, Zeyu Shen, Meng Zhang, Yusheng Shen, Tiantian Li, Casper Ho Yin Chung, Lijuan Zhang, Ji Wang, Binbin Cui, Peng Fei, Yusong Guo, Shengwang Du, and Shuhuai Yao. Single-Frame Deep-Learning Super-Resolution Microscopy for Intracellular Dynamics Imaging. *Nature Communications*, 14(1):2854, May 2023. 1

[6] Alex Clark. Pillow (pil fork) documentation. 2015. 5

[7] 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(1):12050, Aug. 2019. 1

[8] Terrance DeVries and Graham W. Taylor. Improved Regularization of Convolutional Neural Networks with Cutout. arXiv:1708.04552, Nov. 2017. 2

[9] Claude E. Duchon. Lanczos Filtering in One and Two Dimensions. *Journal of Applied Meteorology and Climatology*, 18(8):1016–1022, Aug. 1979. 5

[10] Jeffrey M. Ede. Deep Learning Supersampled Scanning Transmission Electron Microscopy. arXiv:1910.10467, Oct. 2019. 3

[11] Jeffrey M Ede. Warwick Electron Microscopy Datasets. *Machine Learning: Science and Technology*, 1(4):045003, Dec. 2020. 2, 3, 7, 8

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

[13] Li Fan, Zelin Wang, Yuxiang Lu, and Jianguang Zhou. An Adversarial Learning Approach for Super-Resolution Enhancement Based on AgCl@Ag Nanoparticles in Scanning Electron Microscopy Images. *Nanomaterials*, 11(12):3305, Dec. 2021. 1

[14] 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, 2

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

[16] Katsumi Hagita, Takeshi Higuchi, and Hiroshi Jinnai. Super-Resolution for Asymmetric Resolution of FIB-SEM 3D Imaging using AI with Deep Learning. *Scientific Reports*, 8(1):5877, Apr. 2018. 1

[17] Richard Wesley Hamming. *Digital Filters 2nd ed.* Prentice-Hall, 1983. 5

[18] Dan Hendrycks, Andy Zou, Mantas Mazeika, Leonard Tang, Bo Li, Dawn Song, and Jacob Steinhardt. PixMix: Dreamlike Pictures Comprehensively Improve Safety Measures. In *Proc. of CVPR*, pages 16762–16771, New Orleans, LA, USA, June 2022. 2, 3, 4, 5, 6, 7, 8

[19] 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. 6

[20] Björn Möller, Jan Pirklbauer, Marvin Klingner, Peer Kasten, Markus Etzkorn, Tim J. Seifert, Uta Schlickum, and Tim Fingscheidt. A Super-Resolution Training Paradigm Based on Low-Resolution Data Only To Surpass the Technical Limits of STEM and STM Microscopy. In *Proc. of CVPR Workshops*, pages 4262–4271, Vancouver, BC, Canada, June 2023. 1, 2, 4, 5, 6

[21] 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. 5

[22] Chang Qiao, Di Li, Yuting Guo, Chong Liu, Tao Jiang, Qionghai Dai, and Dong Li. Evaluation and Development of Deep Neural Networks for Image Super-Resolution in Optical Microscopy. *Nature Methods*, 18(2):194–202, Feb. 2021. 1

[23] David Salomon. *Data Compression: The Complete Reference*. Springer Science & Business Media, 2004. 6

[24] 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

[25] Amit Suveer, Anindya Gupta, Gustaf Kylberg, and Ida-Maria Sintorn. Super-Resolution Reconstruction of Transmission Electron Microscopy Images Using Deep Learning. In *16th International Symposium on Biomedical Imaging*, pages 548–551, Apr. 2019. 1

[26] Radu Timofte, Rasmus Rothe, and Luc Van Gool. Seven Ways to Improve Example-Based Single Image Super Resolution. In *Proc. of CVPR*, pages 1865–1873, Las Vegas, NV, June 2016. 2

[27] 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 6000–6010, Long Beach, CA, USA, Dec. 2017. 6

[28] Jia Wang, Chuwen Lan, Caiyong Wang, and Zehua Gao. Deep Learning Super-Resolution Electron Microscopy Based on Deep Residual Attention Network. *International Journal of Imaging Systems and Technology*, 31(4):2158–2169, 2021. 2

[29] 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. 6

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

[31] Jaejun Yoo, Namhyuk Ahn, and Kyung-Ah Sohn. Rethinking Data Augmentation for Image Super-Resolution: A Comprehensive Analysis and a New Strategy. In *Proc. of CVPR*, pages 8372–8381, Seattle, WA, USA, June 2020. 2

[32] Sangdoo Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, and Youngjoon Yoo. CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features. In *Proc. of ICCV*, Aug. 2019. 2

[33] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, and David Lopez-Paz. Mixup: Beyond Empirical Risk Minimization. arXiv:1710.09412v2, Apr. 2018. 2