Towards Scanning Electron Microscopy Image Denoising: A State-of-the-art Overview, Benchmark, Taxonomies, and Future Direction

Sheikh Shah Mohammad Motiur Rahman[1*], Michel Salomon[1†] and Sounkalo Dembélé[1†]

[1*]FEMTO-ST Institute, UMR 6174 CNRS, Université de Franche-Comté, Belfort and Besançon, F-90000 and F-25000, France.

*Corresponding author(s). E-mail(s): sheikh.rahman@femto-st.fr;
Contributing authors: michel.salomon@femto-st.fr;
sounkalo.dembele@femto-st.fr;
†These authors contributed equally to this work.

Abstract

Purpose: A scanning electron microscope (SEM) enables imaging of micro scale objects. It is an analytical tool widely used material, earth and life sciences. However, SEM images often suffer from high noise levels, influenced by factors like dwell time, the time during which the electron beam stays per pixel at acquisition. Slower dwell times reduce noise but risk damaging the sample, while faster ones introduce uncertainty. To address this, state-of-the-art (SOTA) denoising techniques need exploration. Experimentation is crucial to identify the most effective methods that balance noise reduction and sample preservation, ensuring high-quality SEM images with enhanced clarity and accuracy. **Methods:** We conducted a thorough analysis tracing the evolution of image denoising techniques, ranging from classical methods to deep learning (DL) approaches. A comprehensive taxonomy of this reverse problem solutions was established, detailing the developmental flow of these methods. Subsequently, we identified and reviewed the latest state-of-the-art techniques based on their reproducibility and the public availability of

their source code. The selected techniques were then tested and investigated using scanning electron microscope images. **Results:** After a thorough analysis and benchmarking, it is evident that existing deep learning based denoising techniques fall short in maintaining a balance between noise reduction and preserving crucial information for SEM images. Issues like information removal and over-smoothing have been identified. To address these constraints, there is a critical need for the development of SEM image denoising techniques that prioritize both noise reduction and information preservation. Additionally, it is found that the combination of multiple networks, like generative adversarial network (GAN) and convolutional neural network (CNN), known as BoostNet, or vision transformer (ViT) and CNN, known as SCUNet, improves denoising performances. **Conclusion:** It is recommended to use blind techniques to denoise real noise while taking into account detail preservation and tackling excessive smoothing, particularly in the context of SEM. In the future the use of explainable AI will facilitate the debugging and the identification of these problems.

Keywords: Scanning electron microscopy (SEM), Image denoising, State-of-the-art (SOTA), Spatial domain denoising, Frequency domain denoising, Deep learning based denoising, Benchmarking

1 Introduction

Digital image processing is now available in a variety of fields, including physics, military sector, medicine, industrial applications, robotics, intelligent transportation systems, and so on [1]. The fields of application clearly show that it is used for critical and sensitive tasks [2].

However, images are inevitably contaminated by noise during acquisition, compression, and transmission due to the effect of the environment, transmission channel, and other variables, resulting in distortion and loss of information in image. Image processing activities, such as video processing, image analysis, and tracking, are harmed by the presence of noise. Thus, work must be done to minimize noise without sacrificing the image quality (edges, corners, and other sharp structures). This is why image denoising is a crucial topic in today's image processing systems [3, 4].

The main mathematical representation of an image with noise is as followed:

$$noisy\_image = clean\_image + noise$$

where the type of noise can vary depending on the capturing device, camera, smartphone, microscope, etc. On the one hand, photon shot noise, fixed pattern noise, dark current, readout noise, and quantization noise are some of the principal sources of real-world noise [5]. On the other hand, there are various types of noise such as Gaussian [6, 7], Poisson [8], Pepper & Salt [9] and so on.

The process of removing noise from a noisy image and restoring the original image is referred to as image denoising. Noise, edges and textures are all high-frequency components making them difficult to distinguish. As a result, denoised images will definitely lose some detail. The ultimate objectives are to smooth flat regions, highlight edges, preserve textures and avoid introducing artifacts. [4]. Image denoising is, in reality, a well-known problem that has been studied for a long time. However, it remains a difficult and unfinished process. The major reason for this is because it is an inverse problem with no unique solution from a mathematical standpoint. This is why a wide variety of image denoising solutions have been proposed in the literature. They range from classical mathematical approaches to modern deep learning, via simple networks, convolutional neural networks (CNNs), transformers, generative adversarial networks (GANs) and other complex networks. It is also noticeable that the attention comes with the consideration of various noise sources, different types of noise, as well as blind denoising.

This study focuses on a detailed review, taxonomy, and benchmark of different approaches proposed in the domain. In addition, it opens a dimension of denoising in the area of scanning electron microscopy (SEM) with future direction. In fact, the images from SEM contain a lot of noise, especially when the scanning speed is fast. Therefore, this study will start by describing the evolution of image denoising techniques, from a taxonomy to a practical reference of state-of-the-art techniques. The aim of this paper is to study, compare and evaluate the performance of different approaches from various fields in the application of SEM image denoising, which may lead us to open up new direction for further research.

The major contributions of this paper are as follows:

- a survey showing the evolution of image denoising techniques, from mathematics to deep learning;
- a detailed taxonomy of image denoising techniques wherever possible;
- a detailed benchmark of different techniques on different areas in one place which may help others to decide to start with;
- a new direction on SEM images processing;
- a comprehensive range of state-of-the-art techniques, with assessments of the latest and most effective approaches.

The remaining sections of this paper are structured as follows. In Section 2, we present the evoluation of image denoising techniques and develop a taxonomy, with a focus on the latest deep learning models. Section 3 benchmarks several state-of-the-art techniques. In Section 4, a detailed discussion of the results provided by the benchmarking is carried out. Finally, Section 5 concludes the paper with some future directions.

2 Evolution of image denoising: a taxonomy

The evolution of image denoising techniques ranges from spatial domain to neural network via transform i.e., frequency domain. Thus, the first level of the taxonomy, the type of techniques, is composed of three classes as shown in Figure 1. Each class can then be further decomposed as shown in Figures 2, 3, and 4, respectively.

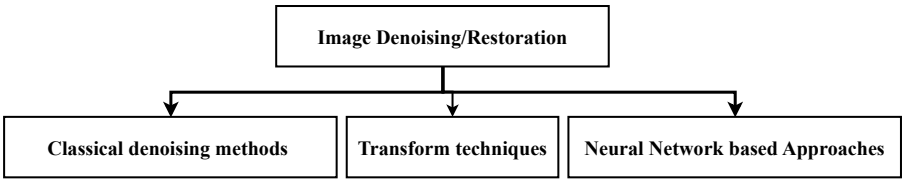


Fig. 1: Types of image denoising techniques

2.1 Classical techniques - Spatial domain

A long-established denoising approach is spatial domain approach. It consists in applying spatial filters directly to images [10]. The two types of spatial domain techniques are spatial filtering and variational denoising techniques [4] as shown in Figure 2.

Spatial filtering is low-pass filtering that presupposes the presence of noise in a higher frequency range. Normally, spatial filters reduce noise to a decent degree but at the expense of image blurring, which results in the loss of sharp edges [11]. Interpolation and resampling were among the various preprocessing and other operations that were previously carried out using mean filtering, followed by Wiener filtering. Usually non-linear filtering [12] works better than linear one. It includes median filtering, weighted median filtering or bilateral filtering [13].

In variational denoising, the denoised image is calculated by minimizing an energy function. Finding an appropriate image prior is crucial for this approach. Gradient priors, non-local self-similarity (NSS) priors, sparse priors, and low-rank priors have all shown to be successful prior models. Total variation (TV) with regularization [14–16], anisotropic diffusion-based method [17, 18] and fast gradient-based method [19] are proposed to solve the disadvantages of over-smooths [20], staircase effect in flat areas and so on [21]. Later on, the performances of these local denoising methods with high noise were found low. After that, NSS-prior [22], non-local means (NLM) [23] come out with noticeable improvements. However, the improvement of NLM along with the combination of TV regularization has been proposed [24]. It outperformed the other techniques in terms of noise suppression, but not in terms of visual quality or structural information in the images. Moreover, based on figuring out how often image patches [25] are and using weighted nuclear norm

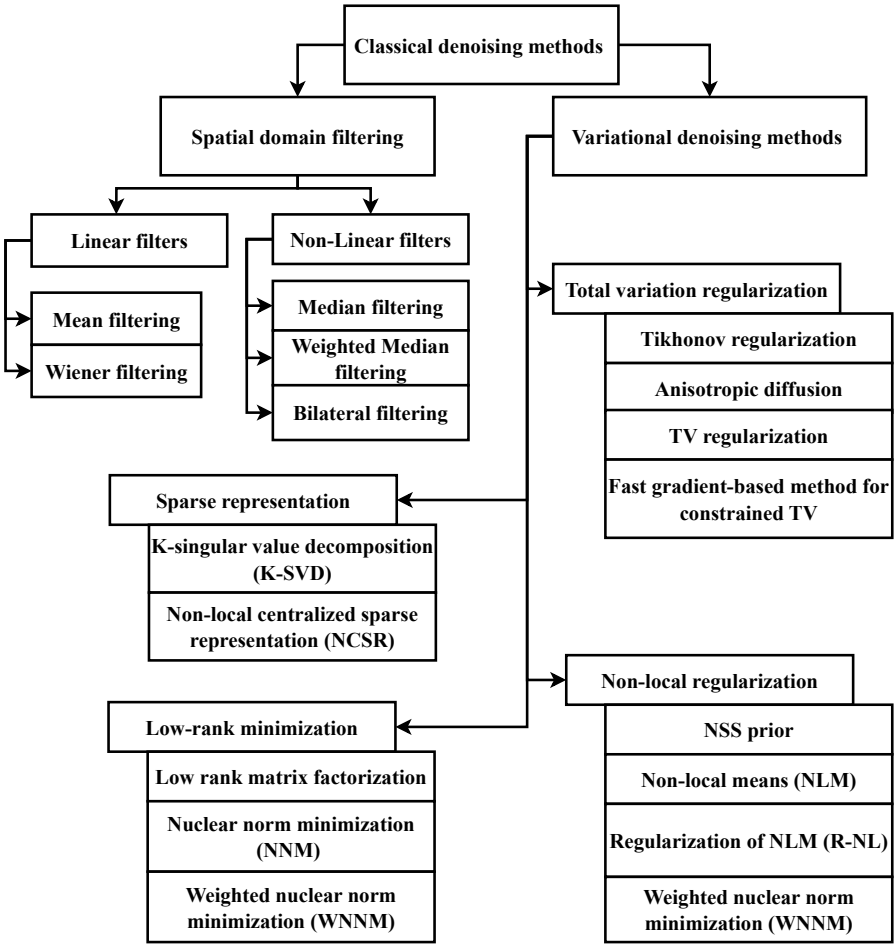


Fig. 2: Spatial domain techniques

minimization - WNNM to take advantage of the low-rank attribute, a great improvement and extension of NSS techniques was proposed [26, 27].

The sparsity prior of natural images is used by several contemporary image denoising algorithms including K-singular value decomposition (K-SVD) algorithm [28, 29] and non-local centralized sparse representation (NCSR) [30] which is coupled with NSS prior. Apart from the sparsity representation of images patches, a matrix representation of the same patches is done by low-rank minimization. The methods by low-rank minimization are in two groups named low rank matrix factorization and nuclear norm minimization (NNM). Despite the fact that most low-rank minimization approaches (particularly the WNNM method) outperform earlier denoising methods, the iterative boosting step has a significant computing cost [4].

2.2 Transform techniques - Frequency domain

Transform techniques started with the Fourier transform but then gradually developed with cosine transform, wavelet methods [31], block-matching and 3D filtering (BM3D) [32] (see Figure 3). Transform techniques focused on transforming the noisy images to another domain and then denoise based on the different features of the transformed images. During transformation, two transformation functions can be used, namely the data-adaptive function and the non-data-adaptive function. [33]. However, BM3D is also a popular method of collaborative filtering, which is an extension of non local method in transform domain [4].

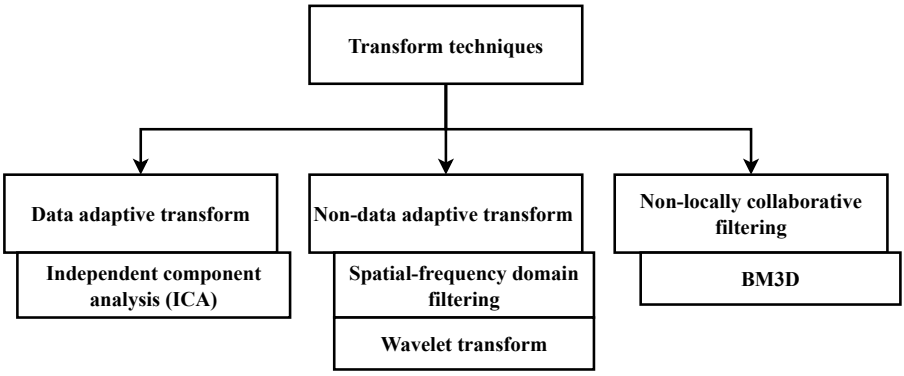


Fig. 3: Transform domain techniques

Principal Component Analysis - PCA [34, 35] and Independent component analysis - ICA [36, 37] are data adaptive methods. But those methods need a noise free sample which may not be possible for all the time in practice. On the other hand, spatial frequency and wavelet transform [38] are counted as non-adaptive data methods. Spatial frequency takes a long time and is dependent on the cut-off frequency as well as the behavior of filter function. The wavelet transform, on the other hand, is significantly reliant on the wavelet bases used. If the selection is incorrect, the images displayed in the wavelet domain will not be accurately represented, resulting in a poor denoising impact [4].

Back to the BM3D technique as it was mostly used and most popular technique once, there have been many attempts at improvement, as performance is sometimes mediocre, especially at high noise levels. This means that as noise increases, BM3D performance decreases mainly in flat areas. Thus, a new dimension of image denoising techniques is launched with neural network-based approaches.

2.3 Neural network based techniques

2.3.1 From plain neural networks to deep neural networks

The neural networks-based approaches have their own history of evolution in image denoising which started with plain neural network or encoder-decoder or multilayer perceptron (MLP). Encoder-Decoder, for example RED-Net [39] can be convolutional neural network (CNN) or recurrent encoder-decoder network, where MLP can be categorised as feedforward artificial neural network (ANN). However, initially all these techniques were based on supervised CNN, but this changed over the time. In this section, we will give details on these techniques, including the network architectures used (see Figure 4).

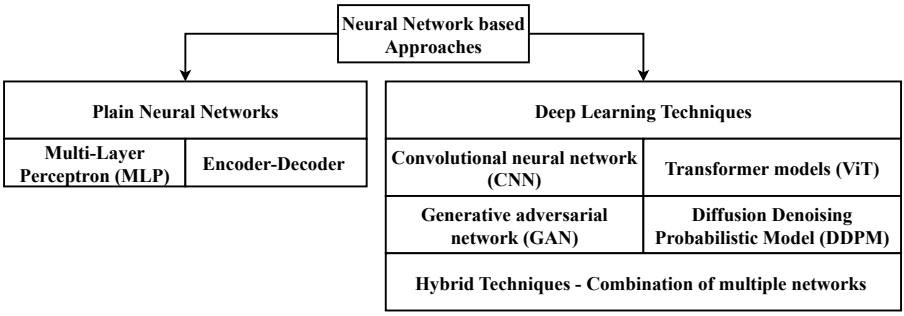


Fig. 4: Neural network-based techniques

Plain MLP as plain neural networks were applied in image patches, and compted with the conventional techniques in denoising Gaussian noise [40]. Residual Encoder-Decoder Network (RED-NET) [41] based techniques were also investigated for image denoising in [42–44]. However, deep neural networks (DNN) or deep learning techniques are the main attraction with remarkable performances from last several years. Depending on the type of network, it may be a Convolutional Neural Network (CNN), a Vision Transformer (ViT) or a Generative Adversarial Network (GAN).

The initial step in neural network-based image denoising technique was proposed in [45], comparing with FoE (Fields of experts) [46], with a very limited layer. In [47], MLP was applied, which was one of the first to be competitive with the BM3D baseline approach [48]. After that, CSF [49], SRCNN [50], ARCNN [51], and TNRD [52] were proposed with comparable performances with the other techniques. With the gradually growth of deep neural networks, the types of learning processes started to get noticed by researchers, because, it is needed in supervised techniques to have a clean target image for training which may not be possible in practice. Thus, it became an emerging issue to investigate further.

There are four ways of learning process in state-of-the-art techniques: supervised, unsupervised, semi-supervised, and self-supervised. The presence

of labels in the training data subset distinguishes the supervised, unsupervised and semi-supervised technique groups. Along with the usage of input features, supervised techniques incorporate a predefined output attribute (label) [53, 54]. Conversely, training without the use of a target label is referred to as unsupervised learning [54]. Semi-supervised learning solved an existing problem, namely having a large number of labeled data or discovering unlabeled data by mixing an unknown label with a known label. Thus, learning from a large amount of unlabelled data incorporated with labeled data is defined as semi-supervised learning process [55]. Self-supervised learning is a learning process that depends on pretext tasks that may be formed using solely unsupervised data. A pretext assignment is designed so that accomplishing it necessitates the acquisition of a useful visual representation [56].

Most of the neural network-based image denoising proposed frameworks are based on CNN, including DnCNN in 2017 [57], FFDNet in 2018 [58], ECND-Net [59] and SDNet in 2019 [60], BRDNet [61] and ADNet [62] in 2020. All these mentioned methods are based on supervised learning framework which needs clean images as target to train. But in practice it may not be possible all the time. Thus, the attention on supervised learning based denoising decreased gradually. Over the time, unsupervised, semi-supervised and self-supervised frameworks are getting attention to investigate in image denoising. Moreover, not only CNNs but also transformers (ViT) and GANs provide a more promising role in image denoising.

2.3.2 Deep neural networks

Deep neural networks (DNN), or deep learning techniques, are promising and the most successful methods nowadays. Thus, we will focus on these techniques with the explanation of maximum possible variations and the evolution of these techniques (see Figure 4).

Convolutional neural networks (CNN)

This section will focus on the techniques with CNN architecture but with unsupervised, semi-supervised and self-supervised learning schemes, which have been proposed in recent years.

With the domination of deep neural networks in image denoising, it was needed to have clean and noisy images in supervised learning. But recently, training without clean target has been proposed and it was found that it is possible to train without clean target. Without clean target, i.e. ground truth, only with pairs of noisy images, Noise2Noise (N2N) [63] enabled the training of CNN in 2018. Furthermore, N2N outperformed BM3D and DIP (Deep Image Prior) [64].

One step further of N2N, based on the motivation of the absence of noisy image pairs, Noise2Void (N2V) [65] performed without noisy pairs as well as clean targets. N2V used the network of U-Net [66] as starting point of implementation inspired from CARE framework proposed for fluorescence

microscopy [67]. N2V enables training from one single noisy image by extracting a patch as input and uses the center pixel of that as target. For achieving their goals, the authors introduced a special network with a receptive field which has a blind-spot in the center of it. Apart from the input pixel at its exact position, the CNN prediction for a pixel is influenced by all input pixels in a square region which defined as blind-spot network by the authors. However, building such a network in place that can nonetheless operate effectively is not easy. They suggested a masking approach to circumvent this issue and produced the same results with any regular CNN: the value in the middle of each input patch is replaced with a randomly chosen value from the surrounding area. This basically wipes the pixel's data and stops the network from learning its identification [65]. However, the randomly chosen values from the surrounding area and the utilized random function are not truly J-invariant (independent feature dimensions function). In addressing this issue, Noise2Self (N2S)[68] was developed under the assumption that noise is statistically independent across multiple measurement dimensions, while the true signals are correlated.

N2S [68] is a self-supervision based system for blind image denoising. It is a framework for denoising high-dimensional measurements that does not rely on prior signal knowledge, noise estimation, or clean training data. To anticipate each other, the characteristics exhibiting a conditionally valid signal were employed, with the condition being independent. That enables N2S to train from a single noisy measurement with performance comparable to that of supervised learning. They suggested that the same method may be possible to use for calibrating other traditional methods like NLM and median filters including for denoising the under-sampled single-cell gene expression data.

In N2V [65], authors claimed that supervised like performance were not possible with the self-supervised techniques. Self-supervised techniques assumed that noise is pixel-wise independent and that the true intensity of a pixel can be predicted from local image context, hence excluding the previously mentioned blind-spots. However, for a variety of applications, including microscopy images, the assumptions may not be fully satisfied and there may be room for improvement [69].

Even in [69], the competence with supervised learning by self-supervised was noticed. They tried to predict Gaussian intensity distributions (per pixel) from assuming or guessing a Gaussian noise model. Similar to [69], Probabilistic Noise2Void (PN2V) [70] has been proposed which is not limited to any noise model like Gaussian noise model or the intensity of prediction. It introduced a way how to leverage information about blind-spots of the network. PN2V is a complete probabilistic model which is now not limited to the selection of statistical estimator to employ. PN2V with an additional noise model was able to improve the results in image reconstruction which was a drawback of previous N2V and N2S approaches. The noise model of PN2V needs to acquire the calibration data. Thus, [71] proposed an improved version of PN2V which is fully unsupervised by replacing the noise model (histogram based) by other noise

models (parametric). They also showed that it is possible to create a noise model which is suitable, even if calibration data is absent. Another study [72] showed that it is possible to improve the blind spot denoising used in N2V and N2S by incorporating additional knowledge from the signal itself which is oversampled and diffraction-limited.

In contrast of blind spot denoising, Noiser2Noise [73] and Noise-As-Clean [74] follow the similar concept to generate synthetic noise from corrupted or noisy images noise model for training pairs. However, there exists a reliance on recognizing a noise model that proves challenging to identify in the real world [75]. Thus, [75] proposed another self-supervised approach named as Neighbor2Neighbor (Nr2N). The training mechanism in Nr2N consists of two stage. In first stage, they generate noisy image pairs from random neighbor sub-samplers. In second stage, they adopt a regularized loss to solve the non-zero ground truth gap issues between the sub-sampled noisy image pairs.

In parallel, recorrupted-to-recorrupted (R2R) [76], an unsupervised technique, was proposed to tackle the overfitting issue which may occur for the absence of clean images. One of the main contribution from R2R was training from unorganized noisy images where the truth targets or the pairs of noisy images were not needed. With blind spot techniques, it is possible to lose information when removing pixels from the image, whereas R2R worked with all image pixels but with little bit higher noise level. In relation to that, there are some techniques which also considered overfitting, for example DIP [64] which used early-stopping and Self2Self (S2S) [77] which used dropout layer based scheme. On the other hand, Stein's Unbiased Risk Estimator (SURE) [78] based techniques, like Net-SURE [79], penalize the prediction divergence in order to regularize the deep neural network.

From N2N to N2V, PN2V, N2S, S2S, and R2R, on the other hand, take a long time to train which is unsuitable for use on high-resolution microscope images in time-sensitive scenarios. To make a fast denoising, Noise2Fast [80] proposed a network which is based on mapping between adjacent pixels as Nr2N [75]. The authors tried to tune the network by using a discrete training set concept of four images and training them on a small network but validating with full sized original images. The discrete training set obtained from a form of downsampling, they named it checker-board downsampling. Authors claimed that they reached very competitive results compared to other SOTA techniques in terms of time as well as performance except S2S. However, S2S takes over 100 times for denoising only a single image [80]. When it comes to approximating neighboring pixels, certain methods based on proximity can lead to excessive smoothing, occasionally resulting in the destruction of structural continuity due to sub-sampling. Recently, the authors of Blind2Unblind [81], again focusing on blindspot based denoising, claimed that the probability of information loss by previous methods can be minimized by using their re-visible loss concept. Nevertheless, their proposal involves implementing a mask mapper to expedite model training and introducing a re-visibility loss.

This re-visibility loss aims to unveil blind spots, enabling the model to directly learn from raw images.

Above mentioned techniques are mostly following self-supervised learning process, but there are a few semi-supervised attempts as well. For example, Noise2Atom, an unsupervised technique, [82] was developed by targeting microscopy images (scanning transmission electron) based on the semi-supervised Multi-scale Convolutional Neural Network (MCNN), and in many cases performs better. One of the most semi-supervised successful approach in image denoising and reconstruction is Noise2Recon [83] that performs joint MRI reconstruction and denoising.

In contrast, CNN-based approach has received remarkable attention since last few years. Apart from above mentioned techniques, there are many other contributions coming from the researchers, including Noise2Same [84], Noise2Inverse [85], Noise2Score [86], MIRNet [87], DudeNet [88], DeepRED [89], DCDicL [90], NBNet [91], Clean-to-Noise (C2N) [92] and so on. Recently, a few other techniques have emerged, such as IDEANet (CNN + GAN based) [93], NSTBNet [94], NFCNN [95], TSLR [96], MalleConv [97], CVF-SID [98] and there are a lot more still adding to the bucket. Furthermore, focusing of sRGB coloring space for digital images CVF-SID [98] proposed a cyclic multivariate function based image denoising method for the real world sRGB images which can work without any prior assumption on noise distribution. However, there is another technique ADL [99] based on transfer learning which is trained to solve one problem, but applied to a different related problem. For example ADL was trained on an object detection dataset but applied to perform Magnetic Resonance Image (MRI) denoising.

Vision transformer models (ViT)

Transformer-based network architectures, unlike CNNs, are naturally effective at capturing long-range relationships in data by using global self-attention. Transformer's achievement in the realm of natural language processing also encourages computer vision researchers [100].

The Transformer, initially introduced by [101], is primarily based on attention mechanisms and has recently found applications in tasks related to image denoising. Although it was initially developed for language translation in Natural Language Processing, its usage has expanded to the field of computer vision, as inspired by [102]. This section will delve into the contributions, as far as our knowledge extends, regarding the utilization of transformers in image denoising.

Self-attention mechanism based technique in image super-resolution started from [103], where a standard blocks of transformer was applied to image processing (Image Processing Transformer or IPT) [104] within a multi-task learning framework. Although, for good performance, IPT depends on multi-task learning and pretraining on a large-scale synthetic dataset, Uformer [100], a U-shaped transformer provides efficient results. Uformer designed two cores named locally-enhanced window (LeWin) and multi-scale spatial

bias. LeWin helps to reduce the computational complexity in the presence of high-resolution feature maps. On the other hand, the learnable multi-scale restoration modulator adjusts features of the decoder in different layers.

A concurrent work with Uformer, SwinIR [105] based on Swin Transformer [106] was proposed. It has three parts including shallow feature extraction (a convolution layer), deep feature extraction (main residual Swin Transformer blocks) and high-quality image reconstruction (reconstruction module). It introduced an approach which worked with diverse application such as image denoising considering grey and color images both, image super-resolution, and JPEG compression artifact reduction.

Motivated by Uformer, Eformer [107] is proposed with Sobel filters by targeting edge enhancement in medical imaging which was further extended with residual learning scheme (Eformer-residual). Eformer-residual outperformed in medical image denoising.

One study [108] proposed TED-Net, a vision transformer-based Encoder-decoder Dilation Network focused on low-dose computed tomography image denoising. TED-Net is a completely convolution free approach, but there are few Convolutional Vision Transformer (CvT) approaches as well [109, 110].

With the help of transformers, recently variety of techniques mostly focusing on restoration have been proposed. For example, SUNet (Swin Transformer with UNet) [111], STFNet (transformer fusion network) [112], DenSformer (based on Uformer and Eformer) [113], Restormer (based on multi-head attention and feed-forward network) [114], CTformer (tokenization or detokenization blocks and a residual encoder-decoder structure) [115] and a lot more, transformers are gaining significant attention and trending for exploration nowadays. However, it is noticeable that most of the transformer-based approaches are focusing image restoration as well as denoising in medical imaging (for example CT and MRI images).

Generative adversarial networks (GAN)

Low-dimensional latent vectors are transformed into aesthetically realistic images using generative adversarial networks (GANs) [116, 117], which are adapted to the purpose of generative modeling. For image classification and object recognition (discriminative learning based issues), GANs surpassed all other previous computer vision algorithms.

There are two networks in GANs. The first one, called generator, actually produces an image from a low-dimensional latent vector. The other component, known as the discriminator, attempts to identify the generated image as a fake sample, thereby attempting to deceive the generator.

However, from the concept of the reverse mapping (from image space to latent space) [118, 119], one study showed that it is possible to denoise images in practice using latent vector recovery process of GAN [117]. It have implemented deep convolutional generative adversarial networks (DCGANs) [120] for the experiments of denoising and showed significant performance, comparable to that of SOTA techniques.

Furthermore, GAN, GAN-CNN, GAN with grouped residual dense network (GRDN) and few more GAN-based approaches are proposed in image denoising [121–133]. Apart from image denoising, super resolution is a close topic which was targeted by few researchers, initially SRGAN [134] then [135–144] and many more.

Most GAN-based denoising methods were not completely pure GAN, but with the mix of other deep learning networks like CNN, RNN, RDN, etc. GAN-based approaches were focusing on image super resolution, reconstruction and restoration. Thus, it is clear from the state-of-art that GAN can be explored further for image denoising specially for low resolution or high noise level images (microscopy images). In addition, most of the studies have used GANs for noise modelling [125] and 3D structure reconstruction [145] which may open a new dimension to work with 3D structure reconstruction after performing an effective denoising that is needed for scanning electron microscopy for 3D metrology of micro-structures.

Diffusion-based deep probabilistic model

Diffusion based model was first introduced in 2015 by [146]. Then in 2019, a different approach but similar concept was introduced in [147]. However, the popular diffusion based approach is Denoising Diffusion Probabilistic Models (DDPM) [148] published in 2020. After that, this kinds of model started to grow faster even with good results.

The diffusion based method is a technique used to create a flexible generative model of data by systematically and gradually altering its structure through an iterative forward diffusion process. After this, a reverse diffusion process is learned to restore the original structure, resulting in a versatile and manageable generative model. This approach enables quick learning, sampling, and probability evaluation in deep generative models. This method employs a Markov chain to gradually transform one probability distribution into another, inspired by principles from non-equilibrium statistical physics [149] and sequential Monte Carlo techniques [150]. Specifically, a generative Markov chain is constructed to convert a known distribution (e.g., Gaussian) into a target data distribution using a diffusion process. Unlike traditional approaches, the probabilistic model is explicitly defined as the endpoint of this Markov chain. Importantly, each step in the diffusion chain has analytically evaluable probabilities, allowing for the analytical evaluation of the entire chain [146].

Over time, this method has overtaken all other techniques in terms of image synthesis, image generation, image super-resolution, image translation, image reconstruction and many others. Recent contributions [151–163] showed the effectiveness of diffusion models in mentioned areas. Especially, diffusion methods outperformed GAN-based models in image generation, denoising and reconstruction. It could therefore be more interesting to study it for denoising scanning electron microscopy images.

Hybrid techniques

Hybrid techniques are usually a combination of multiple network architectures such as "CNN and GAN" or "CNN and ViT". Recently, such kinds of hybrid techniques outperformed single architecture based techniques. For example, BoostNet [164] and SCUNet [165] both exhibit significant performance improvements through the combination of GAN+CNN and ViT(Swin)+CNN(UNet) networks.

3 Benchmark of state-of-the-art techniques

A comparative evaluation of state-of-the-art techniques is carried out in order to select the best architecture to date for image denoising, more specifically, scanning electron microscopy (SEM) images.

One pre-assessment has been performed from the results which are mentioned in the state-of-the-art papers for selecting promising techniques for image image denoising. It has been found that the techniques used variety of datasets and the environment of the execution not same all the time. In addition, the availability of code (reproductibility) has also been taken into account. Next, the main benchmark evaluation was carried out by implementing and evaluating the selected techniques with the microscopy image dataset.

3.1 Dataset information

Most of the state-of-the-art (SOTA) techniques were experimented using existing datasets reported in Figure 5. But we targeted to have the assessment for SEM images dataset which is not exactly found in the literature. There is one SEM dataset which is mainly focused on classification tasks. We therefore selected one of the most critical SEM image samples, the pollen grain, collected by the FEMTO-ST laboratory, for comparative evaluation. However, in the dataset there are different samples based on different dwell times. In fact, we found in a previous work that noise changes in type and level as a function of dwell time [166]. Dwell time represents the duration during which the electron beam engages with the surface atoms of the sample, determining the period required for pixel acquisition and consequently influencing scanning speed. In this study, we used the dataset from the Zeiss Auriga FE SEM including the noise type information found in the paper [166] as shown in Table 1.

In fact, we categorized the scan speed into slow, medium, and fast based on the dwell time range, because we observed that slow speed may damage the samples, while medium and fast speeds do not harm the sample but are prone to higher noise. Therefore, our objective was to test the samples at least at a medium speed to ensure compatibility in practical applications.

3.2 Results in paper of SOTA techniques

Initially, a few techniques have been chosen based on the availability and reproducible capacity of source code as well as the accessibility of pretrained

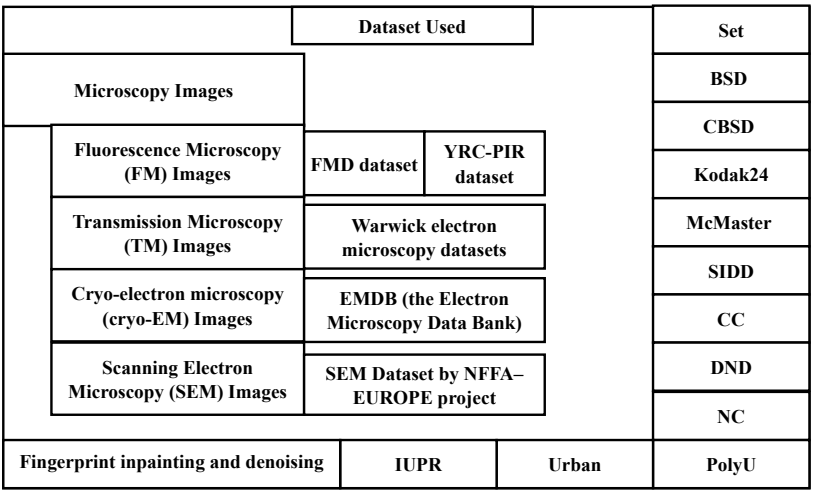


Fig. 5: Existing datasets in the literature

Table 1: Example of Zeiss Auriga SEM configurations

Scan Speed	*Dwell time*	*Noise type*	*Categorization*
8	10 microseconds	Gaussian	Slow
5	2 microseconds	Gaussian	Slow
3	**512 nanoseconds**	**Gamma**	**Medium**
2	280 nanoseconds	Gamma	Medium
1	70 nanoseconds	Gamma	Fast

models provided by the authors of SOTA techniques. Tables 2 and 3 describe the gained results by authors with their respective environments and datasets. In fact, there are a lot of datasets, categorized based on the noise types and sources, in the literature (see Figure 5). Table 4 summarizes, in ascending order of year, the techniques selected for experimental evaluation.

We observed the existence of datasets based on gray-scale and color images for conducting experiments on noise removal. Some techniques performed denoising by removing noise while other techniques are used for image super resolution or up-scaling the resolution. Among all of the datasets, Fluorescence Microscopy Denoising (FMD) dataset is specially for microscopy image denoising which is dedicated to Poisson-Gaussian denoising. It includes around 12000 images collected from real fluorescence microscopy. It is noticed that there is no scanning electron microscopy (SEM) image dataset for denoising task. However, there is one scanning electron microscopy images dataset [175] for classification related tasks which includes 10 types of materials in total.

On the referenced articles, there are the references of mentioned datasets in the Tables 2 and 3. Thus, we did not cite the references of the datasets

Table 2: Results in paper of SOTA denoisers

Techniques	Dataset Info.	Noise Info. / Type	PSNR/SSIM [1]
DnCNN [57]	BSD68 (Gray)	Gaussian ($\sigma = 50$)	26.10 / 0.7076
IRCNN [167]	BSD68 (Color)	Gaussian ($\sigma = 50$)	27.86/NA
FFDNet [58]	BSD68 (Gray)	Gaussian ($\sigma = 50$)	26.29/NA
	BSD68 (Color)		27.96 /NA
	Kodak24		28.98 /NA
	McMaster		29.18 /NA
IMDN [168]	Set5	Up-scaling	38.00 / 0.9605
	Set14		33.63 / 0.9177
	BSD100		32.19 / 0.8996
	Urban100		32.17 / 0.9283
	Manga109		38.88 / 0.9774
Self2Self [77]	Set9	Gaussian ($\sigma = 50$)	29.25/NA
	BSD68		28.70/NA
	BSD68		26.56/NA
	FMD		41.70/0.9725
BoostNet [164]	BSD68(color)	Gaussian ($\sigma = 50$)	28.27/NA
	Kodak24		29.45/NA
	McMaster		29.80/NA
MPRNet [169]	SIDD	Real from smartphone	39.71 / 0.958
	DND		39.80 / 0.954
BSRGAN [170]	DIV2K4D	Gaussian (Super resolution)	24.65/NA

[1]PSNR : Peak Signal-to-Noise Ratio, SSIM : Structural Similarity Method

separately. In fact, our goal is to identify something related to denoising techniques, but for the dataset, we specifically focused on the SEM images dataset if it exists.

3.3 Pre-assessment of SOTA techniques

The initial experimental results obtained from the state-of-the-art (SOTA) techniques are described in Table 4. These results show that no existing model performs better with the pollen grain sample we tested. However, this may be due to the fact that these models have not been trained with samples similar to the pollen grain.

We have noticed during the simulation that self-supervised techniques were taking time to perform denoising. Especially Self2Self [77] took 25000 epochs approximately to reach the level of results described in Table 4 and in Figure 6. In fact, we have targeted SEM images and within minimum time requirements. Thus, self-supervised techniques may not work in our case. In Figure 6, we used an image acquire from SEM with very slow scan speed (scan speed 8) with one iterative denoising by IDR [171] as clean target.

From our initial assessment, we found that BoostNet (CNN+GAN) [164], BSRGAN (GAN) [170], and SCUNet (ViT+CNN) [165] performed better comparatively based on the PSNR / SSIM values and in case of visual evaluation, we found that BSRGAN (GAN) [170] have blur effect problems. It is found that, in our case, GAN alone is not able to produce good quality, but GAN

Table 3: Results in paper of SOTA denoisers (continued)

Techniques	Dataset Info.	Noise Info. / Type	PSNR/SSIM [1]
SwinIR [105]	Train: DIV2K + Flickr2K Flickr2K Test: Set5 Set14 BSD100 Urban100 Manga109	Up-scaling	 38.42 / 0.9623 34.46 / 0.9250 32.53 / 0.9041 33.81 / 0.9427 39.92 / 0.9797
SwinIR+ [105]	Train: DIV2K +Flickr2K Flickr2K Test: Set5 Set14 BSD100 Urban100 Manga109	Up-scaling	 38.46 / 0.9624 34.61 / 0.9260 32.55 / 0.9043 33.95 / 0.9433 40.02 / 0.9800
Noise2Fast [80]	Set12 BSD68 Confocal (FMD)	Gaussian ($\sigma = 50$)	25.82/7.81 25.23/6.70 36.61/9.33
IDR [171]	Kodak BSDS300 BSD64	Gaussian ($\sigma = 50$)	29.27 / 0.803 28.25 / 0.802 26.25 / 0.726
SUNet [111]	CBSD68 Kodak24	Gaussian ($\sigma = 50$)	27.85 / 0.799 29.54 / 0.810
Blind2Unblind [81]	KODAK BSD300 SET14	Gaussian ($\sigma = 50$) Poisson ($\lambda = 30$)	Gaussian: 32.34/0.872 30.86/0.861 31.14/0.857 Poisson: 31.07/0.857 29.92/0.852 30.10/0.844
NAFNet [172]	SIDD	Real from smartphone	39.96 / 0.960
MIRNetV2 [173]	SIDD DND	Real from smartphone	39.84 / 0.959 39.86 / 0.955
Restormer [114]	SIDD DND	Real from smartphone	40.02 / 0.960 40.03 / 0.956
SCUNet [165]	BSD68 (color)	Gaussian ($\sigma = 50$)	29.37/0.8135
Diffusion [174]	SIDD DND	Real from smartphone	41.39 / 0.9744 40.54 / 0.9676

[1]PSNR : Peak Signal-to-Noise Ratio, SSIM : Structural Similarity Index Measure

with CNN or ViT may perform better. Moreover, the techniques employed thus far have not been trained with SEM images, potentially contributing to their weak performance. Therefore, we have chosen to conduct additional experiments, delving into the details of these techniques by training them with SEM images. Our selection of techniques for further experimentation is based on

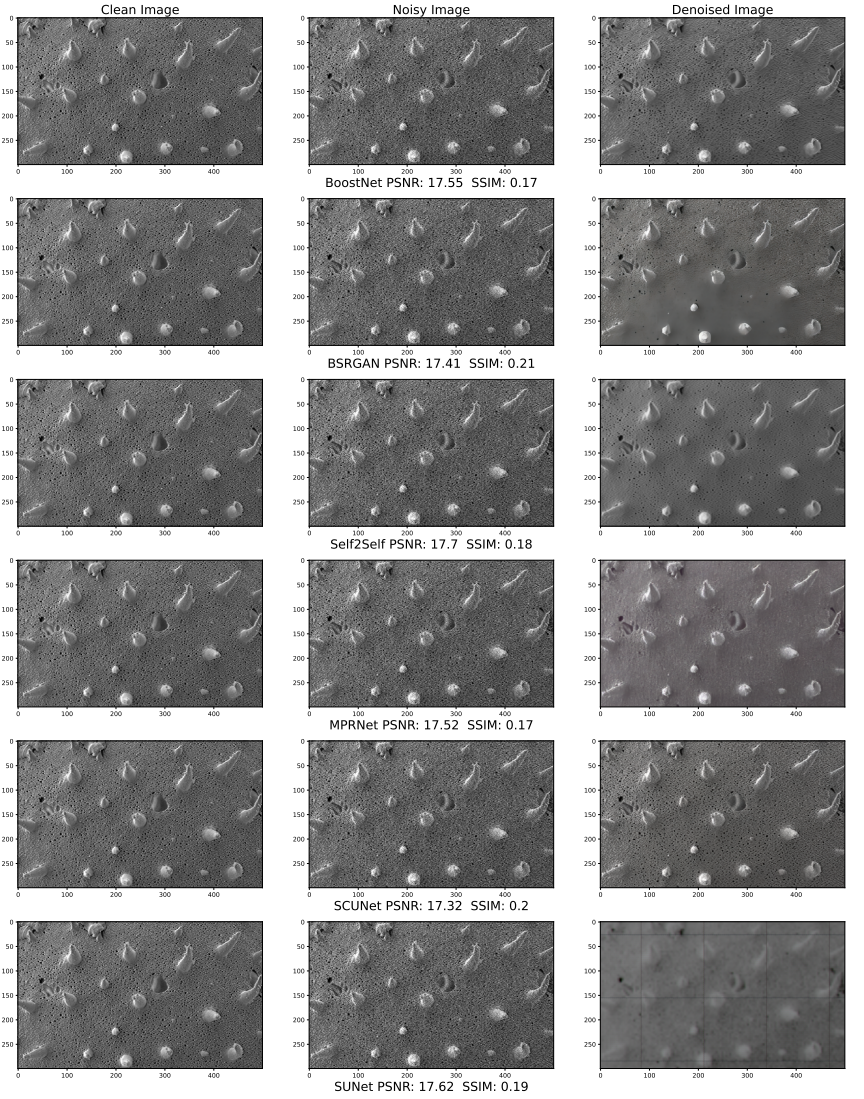


Fig. 6: Visual comparison of SOTA techniques

the availability of training codes and their adaptability to our system. Additionally, several diffusion-based models have demonstrated impressive results in generative modeling and denoising compared to transformers, CNNs, and GANs [174]. We did not make a test with the diffusion model proposed in [174] for denoising real noisy images, because of the unavailability of the codes in public. However, we decided to train and test diffusion models later.

Table 4: Experimental results obtained from SOTA denoisers

Techniques	Year	Data	Learning	Network	PSNR/SSIM [1]
DnCNN [57]	2016	Dataset	Supervised	CNN	16.34/0.09
IRCNN [167]	2017	Dataset	Supervised	CNN	16.25/0.08
FFDNet [58]	2017	Dataset	Supervised	CNN	16.33/0.09
IMDN [168]	2019	Dataset	Supervised	CNN	17.36/0.15
Self2Self [77]	2020	Single	Self-supervised	CNN	17.70/0.18
BoostNet [164]	2021	Dataset	Unsupervised Supervised	GAN[2] CNN	17.55/0.17
MPRNet [169]	2021	Dataset	Supervised	CNN	17.52/0.17
BSRGAN [170]	2021	Dataset	Unsupervised	GAN[2]	17.41/0.21
Real-ESRGAN [176]	2021	Dataset	Unsupervised	GAN[2]	16.63/0.20
SwinIR [105]	2021	Dataset	Supervised	ViT[3]	15.57/0.15
SwinIR+ [105]	2021	Dataset	Supervised	ViT[3]	15.65/0.08
Noise2Fast [80]	2021	Single	Self-supervised	CNN	16.52/0.09
IDR [171]	2022	Single	Self-supervised	CNN	16.33/0.09
SUNet [111]	2022	Dataset	Supervised Transfer	ViT[3]	17.62/0.19
Blind2Unblind [81]	2022	Single	Self-supervised	CNN	16.52/0.09
NAFNet [172]	2022	Dataset	Supervised	CNN	17.65/0.15
MIRNetV2 [173]	2022	Dataset	Supervised	CNN	16.76/0.10
Restormer [114]	2022	Dataset	Supervised	ViT[3]	17.01/0.11
SCUNet [165]	2023	Dataset	Supervised	ViT[3] CNN	17.32/0.20

Note: All techniques were tested and evaluated using pretrained models except the single image denoising.

[1]PSNR : Peak Signal-to-Noise Ratio, SSIM : Structural Similarity Method

[2]Generative Adversarial Networks

[3] Vision Transformer

3.4 Benchmarking of selected SOTA techniques

As described in early assessment, in this stage, we will investigate and benchmark the GAN, ViT and diffusion based model in SEM images denoising. Hence, we intend to create a new microscopy dataset comprising both clean and noisy images. Subsequently, we will train, validate, and test the selected state-of-the-art (SOTA) techniques on this dataset.

Table 5 shows the results of few techniques retrained on a customized SEM image dataset prepared with 500 high quality images (based on manual selection) from [175]. In fact, we selected around 3900 images based on the resolution (Height × Width) programmatic way and then performed the manual selection of 500 images because there were mixed of clean, blurry images in the initial selection. Moreover, we did not carry out all the epochs as conducted by the authors ("In code" values) due to time constraints. We targeted initially to experiment by using 2 days for each training, but the GAN required a training time of 4 days on the Tesla V100 SXM2 32GB GPU. However, after learning we did not achieve improved performance in SEM images, except for

Table 5: Results of some SOTA denoiser experiments (the denoisers are trained with SEM images)

Techniques	Epochs / In code	Time	PSNR / SSIM [1]
DnCNN [57]	10K / 100K	Avg. 2 days	16.66 / 0.10
FFDNet [58]	10K / 100K	Avg. 2 days	16.26 / 0.08
BSRGAN [170]	3.5 K / 200K	Avg. 4 days	17.10 / 0.18
USRNet [177]	5K / 100K	Avg. 2 days	17.53 / 0.15
SCUNet [165]	6.2K / 100K	Avg. 2 days	16.12 / 0.07

[1]PSNR : Peak Signal-to-Noise Ratio, SSIM : Structural Similarity Index Measure
Tested on GV100GL - Tesla V100 SXM2 32GB and noise level $\sigma = 50$

DnCNN [57]. In fact, the authors did experiments with a far more larger number of training epochs. For example, the SCUNet authors used around three days to train on their four NVIDIA RTX 2080 Ti GPUs [165]. Thus, it is concluded that it may be possible to enhance performance in some cases, but would require more training on SEM images with a large number of epochs, or might require hyper-tuning of parameters in SEM image feature engineering, or even might require a different training setup for blind real image denoising. Additionally, we noticed that it is not only difficult but also challenging to remove noises from SEM images using existing techniques as well. We believe that the uncertainty of image noise type and level [166, 178] in SEM images prevents the techniques from performing better, or even to reach at a certain level in SEM images denoising.

4 Discussion and Future Direction

Based on the experiments and benchmarking, it has been found that CNN is the ultimate target network which is needed in image denoising, especially U-Net encoder-decoder network. In fact, GAN alone cannot deliver better performance. We observed that it struggles to preserve features in denoised images, often resulting in over-smoothing effects. On the other hand, ViT alone even cannot performed better in SEM image denoising. However, ViT (Swin-Conv) with the combination of CNN (SCUNet) performed better based on the results found in Table 3 (PSNR/SSIM value of 17.32/0.20) and Figure 6. Eventually, GAN with CNN (BoostNet) also performed better (PSNR/SSIM value of 17.55/0.17), but visually it removes some details from the images. The main problem we found in SOTA techniques is either not working or removing details from the SEM samples.

4.1 Analysis and discussion

In this section, we will make a detailed analysis of the denoised images obtained for our pollen grain sample using the three best SOTA techniques found by the benchmarking. These techniques, BSRGAN-based GAN, BoostNet-based

GAN+CNN and SCUNet-based ViT+CNN, outperformed the other techniques in terms of SEM images. The pollen grain SEM image sample used to conduct the experiments is one of the most critical and complex sample. The visual analysis and findings from different regions of interest (ROI) in the images are depicted in Figures 7, 8, 9, 10, and 11, respectively.

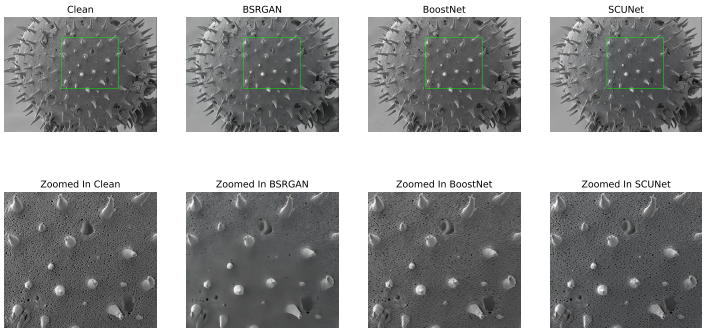


Fig. 7: Visual details for the three best-performing SOTA techniques

Figure 7 represents the larger portion of the images. It can be seen that BSRGAN provides over-smoothed ROI, whereas BoostNet and SCUNet looks quite better. However, in Figure 8, it is clearly identified that in very critical structures, almost all techniques failed to remove the noise. In fact, in this case, SCUNet does not remove details including noise, but other techniques removed few details with some smoothing issues.

Let's move on to the case study shown in Figure 9, which highlights more details. We can observe that all techniques have issues to denoise with exact details like the light white color shape getting changed in the denoised versions. In fact, in Figures 10 and 11, BSRGAN evidently exhibits over-smoothing issues, while the two other techniques share similar details at a quick glance. However, there are subtle pixel differences in the output.

We performed this analysis to go in details of the effects of the three selected SOTA denoisers on a representative SEM image sample. Out of these three techniques visually SCUNet provides promising results with a minor failure. Even, it minimizes the noise with less meaningful information removal. But in terms of PSNR and SSIM, the results are not promising which tends us to go for in-depth insights of the results found from these techniques.

Finally, it is discovered that the objective of denoising without compromising quality and meaningful details is not fulfilled by existing denoisers in the field of SEM images. However, SCUNet introduces a noise synthesis pipeline as a baseline, incorporating various noises such as Gaussian, Poisson, camera

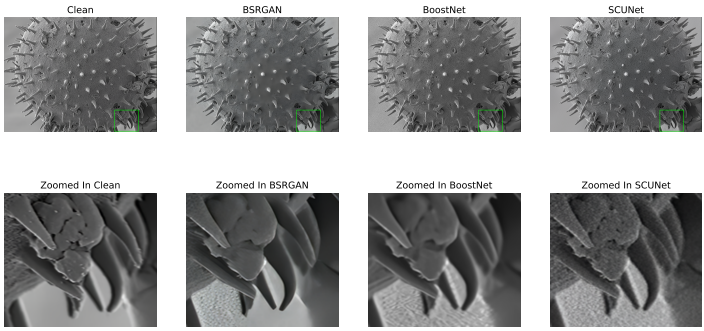


Fig. 8: Visual details for the three best-performing SOTA techniques

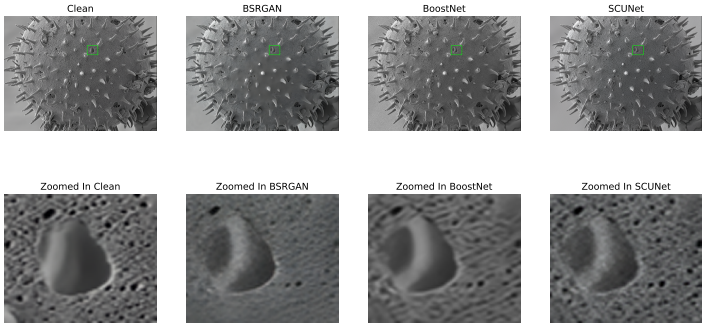


Fig. 9: Visual details for the three best-performing SOTA techniques

noise, and more. This feature positions SCUNet as the leading solution for real blind image denoising, including SEM images

4.2 Features detection and matching

We take one more step in this section to check the maximum features maintained by the SOTA techniques in the images after denoising. Because, if we want to reconstruct the over-smoothed part or the removed details information in the images then at least, we need to have the features or enough texture points in the images. To detect the local features in the images, we used AKAZE [179] local features detector. Regarding the settings, we used a threshold point of 0.0001 and 0.90 for nearest neighbor matching ratio

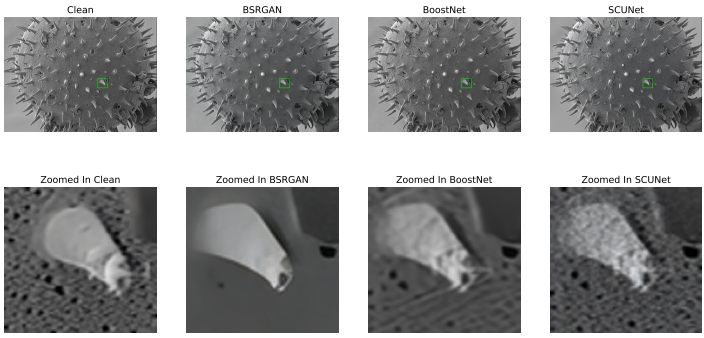


Fig. 10: Visual details for the three best-performing SOTA techniques

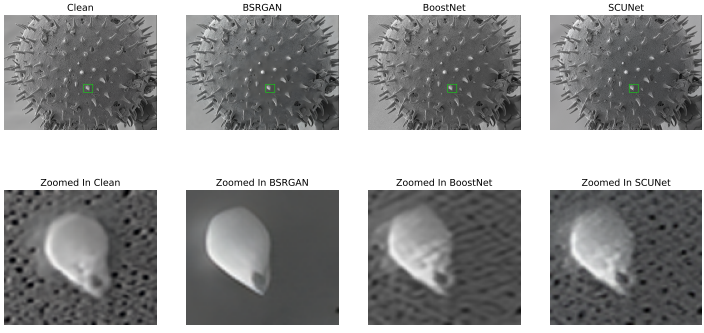
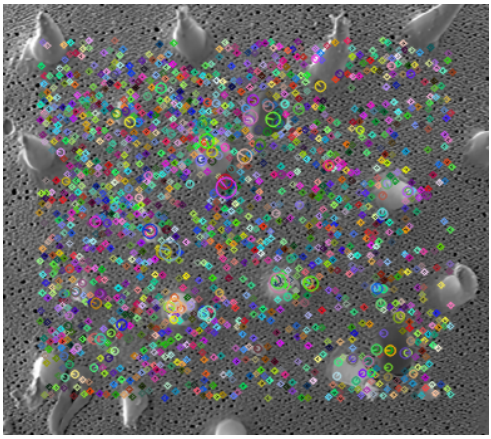


Fig. 11: Visual details for the three best-performing SOTA techniques

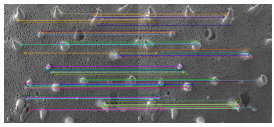
in AKAZE local features binary matching with Brute-force match based on NORM_HAMMING distance.

Figure 12 shows the selected features key-points in the ROI of the clean image (left - (a)) and the matching of those features points with the denoised versions (right - (b), (c) and (d)) of the noisy images. The number of successfully matched features key-points is also noticed, revealing noteworthy insights. SCUNet, in particular, has demonstrated a robust performance by providing a substantial number of matching features, specifically 1290 out of 1621 key-points.

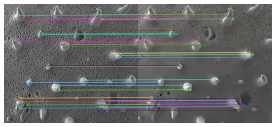
It can be observed that conventional metrics like PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index Measure) fall short in



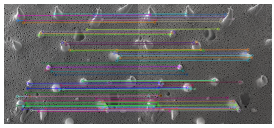
(a) Features key-points in ROI of clean image - 1621 points



(b) Features matched with BoostNet - 1072 points



(c) Features matched with BSRGAN - 322 points



(d) Features matched with SCUNet - 1290 points

Fig. 12: Features detection and matching with AKAZE local features detector of images

adequately assessing SEM image denoising. This leads to the conclusion that these metrics may not be optimal for evaluating the effectiveness of denoising methods in this specific context. In contrast to the lower PSNR and SSIM scores, both BoostNet and SCUNet exhibit superior visual results, effectively preserving meaningful details in the denoised SEM images. This underscores the importance of incorporating visual assessment alongside quantitative metrics in evaluating denoising performance.

Recognizing the persistent challenges in SEM image denoising, there is a suggestion that further improvements can be achieved through post-denoising image reconstruction. This insight indicates an acknowledgment of ongoing efforts and potential advancements in overcoming the existing challenges in this field.

4.3 Comparative Analysis of BoostNet and SCUNet

In this section, we provide a comparative analysis of BoostNet and SCUNet, discussing the reasons behind their superior performance compared to other techniques. Although BoostNet uses a CNN+GAN network architecture and SCUNet a ViT(Swin)+CNN(UNet) one, both have a common target to target "real-world noise model". In fact, the results of SCUNet presented in Table 5 show that it did not perform better because of its training configuration. Indeed, the training configuration mostly depends on the dataset synthesizer.

While BoostNet employs both Gaussian and Poisson noises in the synthesis of noisy images, SCUNet utilizes a distinct training data synthesizer that combines more types of noises. This synthesizer is designed to create real-blind noisy images by incorporating various noises such as Gaussian, Poisson, speckle, JPEG compression, as well as processed camera sensor noises. The processing includes resizing, a random shuffle strategy, and a double degradation strategy. Omitting the utilization of the synthesizer module during our retraining process led to unsatisfactory results, as indicated in Table 5. Thus it can be noticed that the special noise synthesizer including multiple noise types improves the performance in both networks.

We thoroughly experimented with the provided pretrained models of BoostNet and SCUNet and we found that the models which were targeting on "real noise denoising" outperformed. Thus, it is interesting to note that SEM images have a complex local region and very sensitive textures which tend to be excessively smoothed and lose details with other techniques. A priori the proposed or assumed noise model does not perfectly match the actual noise characteristics found in real SEM images.

Compared to BoostNet, SCUNet outperformed due to network selection. BoostNet uses a GAN with CNN where SCUNet uses the combination of DRUNet, SwinIR and UNet architectures as the core of the framework.

BoostNet uses a modified deep residual network based subnetwork to estimate the noise. The subnetwork incorporated into the final model involves signal-dependent noise, encompassing both Poisson and Gaussian noise. A wide range of noise levels [0,75] is used to make it flexible in training. Subsequently, two generators and one discriminator are employed to reconstruct an improved denoised image. Specifically, one generator focuses on denoising, while the other is tasked with constructing details with the assistance of the discriminator to ultimately generate an enhanced output.

Regarding SCUNet, it stands out in the synthesis of real noisy images. The proposed pipeline to synthesize real-noisy images consider various types of noise, alongside resizing. The approach integrates a double degradation strategy and a random shuffle strategy, enhancing its adaptability to diverse challenges and scenarios. This step is followed by the addition of Gaussian noise, Poisson noise, speckle noise, JPEG compression noise, and processed camera sensor noise (including read and shot noise, gamma correction, etc.). This noise addition sequence is repeated twice, and the images are resized. The ultimate goal is to closely approximate real noise in the images, resulting in superior performance compared to other techniques.

We found that SCUNet exhibits superior performance in SEM images, prompting a keen interest in advancing efforts to develop noise synthesis models capable of generating more realistic noises during training. Along with the importance of networks in inverse problems, denoising images obtained by synthesizing real noise types can provide a basis for a uniform blind-real denoising framework.

4.4 Recommendations

From the assessments and this study, we would like to recommend followings to the SEM community to consider when denoising.

- Focusing on real image denoising is essential rather than making blind assumptions about Gaussian noise. As found in our previous study [166], the type and level of noise may change to Gamma or any specific case of Gamma noise.
- It is crucial to concentrate on preserving features from loss, a consideration often overlooked by many techniques. Negative effects on feature preservation can lead to problems, as these features may be mistakenly identified as noise, resulting in over-smoothing and the removal of image details.
- Explainable AI in denoising techniques may aid in understanding the reasons behind blurry or over-smoothed images, as demonstrated in medical imaging by [180].
- Both ViT and CNN (Swin-Conv), as well as GAN and CNN, warrant further exploration, especially in minimizing feature loss or preventing over-smoothing issues.
- While the community is consistently addressing reverse problems, it is recommended, specifically in SEM image denoising, to focus not only on denoising but also on preserving image details.
- Generally, high-resolution images can be achieved through denoising or super-resolution techniques, but SEM images require special attention to preserve significant details with the highest possible resolution.
- Despite details being removed, reconstruction or restoration techniques may be implemented in a second step after denoising to restore meaningful details.
- For complex samples like SEM images, PSNR and SSIM may not be suitable parameters for assessment. Consideration of other metrics such as image key-points is advisable.
- Attention to synthesizing the real noise model is essential, as it has shown significant performance in image denoising.
- It may also be possible to propose a guided network, similar to explainable AI [180], to preserve losses while removing noise from SEM images.

5 Conclusion

Scanning Electron Microscopy (SEM) images are subjected to uncertain changes in noise information based on the setting parameters, dwell time more particularly. Thus, it is important to have a large scale experiment on the state-of-the-art (SOTA) techniques in denoising for those images.

In this work, a vast investigation has been performed to identify the performance of SOTA techniques in SEM images noise denoising. In fact, the paper started with the evaluation of techniques in such reverse problems from classical to intelligent system till date. As a part of this, a taxonomy of these

techniques was presented and discussed in details with the possible timeline of the evolution. A multi-stage benchmarking, as well as comparative experiments, evaluations and justifications were carried out. We have listed the results claimed by the authors in the different articles and carried out experiments with SEM images. Indeed, we evaluated the performance of some pretrained models on other images already available, and finally retrained and evaluated some networks on a dataset of SEM images that we designed.

We found that in inverse problems like image denoising, CNNs mostly have amazing performance, although they also outperform when combined with ViT or GAN. Unfortunately, these models do not work well for SEM images with real noise or when adding synthetic noise, due to the complex structure of the SEM images including the uncertainty of noise type and noise level. They have the problems of removing meaningful important information and over-smoothing. Thus, this study suggests to start rethinking of developing denoising techniques for SEM images by considering the minimization of information removal and over-smoothing problems, particularly focusing on SEM images special characteristics to prevent loss. In this case, explainable AI can contribute to providing transparency to the community regarding the model's success or failure. Finally, we recommend to use blind real noise denoising, emphasizing the preservation of details and addressing over-smoothing, particularly in SEM-like images.

Acknowledgments. This work has been funded by MEB-3D project and supported by the EIPHI Graduate School (contract ANR-17-EURE-0002), partly supported by the French ROBOTEX network (TIRREX ANR-21-ESRE-0015) and its FEMTO-ST technological facility CMNR. We want to acknowledge Kai Zhang, Computer Vision Lab, ETH Zurich, Switzerland for developing the KAIR toolbox [181] which helps us to make our experiments efficiently. Additionally, we used Mesocentre de Calcul de Franche-Comte as GPU server for the experiments.

References

[1] Ya-Lin, S., Chen-Xi, B.: Research and analysis of image processing technologies based on dotnet framework. Physics Procedia **25**, 2131–2137 (2012)

[2] Aslan, M.F., Sabanci, K., Durdu, A.: Comparison of contourlet and time-invariant contourlet transform performance for different types of noises. Balkan Journal of Electrical and Computer Engineering **7**(4), 399–404 (2019)

[3] Jain, P., Tyagi, V.: A survey of edge-preserving image denoising methods. Information Systems Frontiers **18**(1), 159–170 (2016)

[4] Fan, L., Zhang, F., Fan, H., Zhang, C.: Brief review of image denoising techniques. Visual Computing for Industry, Biomedicine, and Art **2**(1),

1–12 (2019)

[5] Xu, J., Li, H., Liang, Z., Zhang, D., Zhang, L.: Real-world noisy image denoising: A new benchmark. arXiv preprint arXiv:1804.02603 (2018)

[6] Bovik, A.C.: The Essential Guide to Image Processing. Elsevier Science, Cambridge, Massachusetts, United States (2009). https://books.google.fr/books?id=6TOUgytafmQC

[7] Liu, W., Lin, W.: Additive white gaussian noise level estimation in svd domain for images. IEEE Transactions on Image processing **22**(3), 872–883 (2012)

[8] Hasinoff, S.W.: In: Ikeuchi, K. (ed.) Photon, Poisson Noise, pp. 608–610. Springer, Boston, MA (2014). https://doi.org/10.1007/978-0-387-31439-6_482

[9] Toh, K.K.V., Ibrahim, H., Mahyuddin, M.N.: Salt-and-pepper noise detection and reduction using fuzzy switching median filter. IEEE Transactions on Consumer Electronics **54**(4), 1956–1961 (2008)

[10] Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Pearson, London, England (2018). https://books.google.fr/books?id=0F05vgAACAAJ

[11] Sanches, J.M., Nascimento, J.C., Marques, J.S.: Medical image noise reduction using the sylvester–lyapunov equation. IEEE transactions on image processing **17**(9), 1522–1539 (2008)

[12] Pitas, I., Venetsanopoulos, A.N.: Nonlinear Digital Filters: Principles and Applications. The Springer International Series in Engineering and Computer Science. Springer, United States (2013). https://books.google.fr/books?id=zWfmBwAAQBAJ

[13] Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In: Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271), pp. 839–846 (1998). IEEE

[14] Willoughby, R.A.: Solutions of ill-posed problems (an tikhonov and vy arsenin). SIAM Review **21**(2), 266 (1979)

[15] Banham, M.R., Katsaggelos, A.K.: Digital image restoration. IEEE signal processing magazine **14**(2), 24–41 (1997)

[16] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D: nonlinear phenomena **60**(1-4), 259–268 (1992)

[17] Perona, P., Malik, J.: Scale-space and edge detection using anisotropic

diffusion. IEEE Transactions on pattern analysis and machine intelligence **12**(7), 629–639 (1990)

[18] Weickert, J.: Anisotropic Diffusion in Image Processing vol. 1. Teubner Stuttgart, Germany (1998)

[19] Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE transactions on image processing **18**(11), 2419–2434 (2009)

[20] Dobson, D.C., Santosa, F.: Recovery of blocky images from noisy and blurred data. SIAM Journal on Applied Mathematics **56**(4), 1181–1198 (1996)

[21] Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of mathematical imaging and vision **40**(1), 120–145 (2011)

[22] Gilboa, G., Osher, S.: Nonlocal operators with applications to image processing. Multiscale Modeling & Simulation **7**(3), 1005–1028 (2009)

[23] Buades, A., Coll, B., Morel, J.-M.: A non-local algorithm for image denoising. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, pp. 60–65 (2005). IEEE

[24] Sutour, C., Deledalle, C.-A., Aujol, J.-F.: Adaptive regularization of the nl-means: Application to image and video denoising. IEEE Transactions on image processing **23**(8), 3506–3521 (2014)

[25] Zoran, D., Weiss, Y.: From learning models of natural image patches to whole image restoration. In: 2011 International Conference on Computer Vision, pp. 479–486 (2011). IEEE

[26] Gu, S., Zhang, L., Zuo, W., Feng, X.: Weighted nuclear norm minimization with application to image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2862–2869 (2014)

[27] Gu, S., Xie, Q., Meng, D., Zuo, W., Feng, X., Zhang, L.: Weighted nuclear norm minimization and its applications to low level vision. International journal of computer vision **121**(2), 183–208 (2017)

[28] Aharon, M., Elad, M., Bruckstein, A.: K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. IEEE Transactions on signal processing **54**(11), 4311–4322 (2006)

[29] Elad, M., Aharon, M.: Image denoising via sparse and redundant representations over learned dictionaries. IEEE Transactions on Image processing **15**(12), 3736–3745 (2006)

[30] Dong, W., Zhang, L., Shi, G., Li, X.: Nonlocally centralized sparse representation for image restoration. IEEE transactions on Image Processing **22**(4), 1620–1630 (2012)

[31] Zhang, L., Bao, P., Wu, X.: Multiscale lmmse-based image denoising with optimal wavelet selection. IEEE Transactions on circuits and systems for video technology **15**(4), 469–481 (2005)

[32] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Transactions on image processing **16**(8), 2080–2095 (2007)

[33] Jain, P., Tyagi, V.: Spatial and frequency domain filters for restoration of noisy images. IETE Journal of Education **54**(2), 108–116 (2013)

[34] Muresan, D.D., Parks, T.W.: Adaptive principal components and image denoising. In: Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429), vol. 1, p. 101 (2003). IEEE

[35] Zhang, L., Dong, W., Zhang, D., Shi, G.: Two-stage image denoising by principal component analysis with local pixel grouping. Pattern recognition **43**(4), 1531–1549 (2010)

[36] Hyvarinen, A., Oja, E., Hoyer, P., Hurri, J.: Image feature extraction by sparse coding and independent component analysis. In: Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No. 98EX170), vol. 2, pp. 1268–1273 (1998). IEEE

[37] Jung, A.: An introduction to a new data analysis tool: Independent component analysis. In: Proceedings of Workshop GK" Nonlinearity"-Regensburg (2001)

[38] Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. IEEE transactions on pattern analysis and machine intelligence **11**(7), 674–693 (1989)

[39] Peng, X., Feris, R.S., Wang, X., Metaxas, D.N.: Red-net: A recurrent encoder–decoder network for video-based face alignment. International Journal of Computer Vision **126**(10), 1103–1119 (2018)

[40] Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising: Can plain neural networks compete with bm3d? In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2392–2399 (2012). IEEE

[41] Mao, X., Shen, C., Yang, Y.-B.: Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. Advances in neural information processing systems **29** (2016)

[42] Couturier, R., Perrot, G., Salomon, M.: Image denoising using a deep encoder-decoder network with skip connections. In: International Conference on Neural Information Processing, pp. 554–565 (2018). Springer

[43] Li, S., Liu, X., Jiang, R., Zhou, F., Chen, Y.: Dilated residual encode–decode networks for image denoising. Journal of Electronic Imaging **27**(6), 063005 (2018)

[44] Li, X., Xiao, J., Zhou, Y., Ye, Y., Lv, N., Wang, X., Wang, S., Gao, S.: Detail retaining convolutional neural network for image denoising. Journal of Visual Communication and Image Representation **71**, 102774 (2020)

[45] Jain, V., Seung, S.: Natural image denoising with convolutional networks. Advances in neural information processing systems **21** (2008)

[46] Roth, S., Black, M.J.: Fields of experts: A framework for learning image priors. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), vol. 2, pp. 860–867 (2005). IEEE

[47] Schuler, C.J., Christopher Burger, H., Harmeling, S., Scholkopf, B.: A machine learning approach for non-blind image deconvolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1067–1074 (2013)

[48] Gu, S., Timofte, R.: A brief review of image denoising algorithms and beyond. Inpainting and Denoising Challenges, 1–21 (2019)

[49] Schmidt, U., Roth, S.: Shrinkage fields for effective image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2774–2781 (2014)

[50] Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence **38**(2), 295–307 (2015)

[51] Dong, C., Deng, Y., Loy, C.C., Tang, X.: Compression artifacts reduction by a deep convolutional network. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 576–584 (2015)

[52] Chen, Y., Pock, T.: Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. IEEE transactions on pattern analysis and machine intelligence **39**(6), 1256–1272 (2016)

[53] Kotsiantis, S.B., Zaharakis, I., Pintelas, P., *et al.*: Supervised machine learning: A review of classification techniques. Emerging artificial intelligence applications in computer engineering **160**(1), 3–24 (2007)

[54] Alloghani, M., Al-Jumeily, D., Mustafina, J., Hussain, A., Aljaaf, A.J.: A systematic review on supervised and unsupervised machine learning algorithms for data science. Supervised and unsupervised learning for data science, 3–21 (2020)

[55] Zhu, X.J.: Semi-supervised learning literature survey (2005)

[56] Zhai, X., Oliver, A., Kolesnikov, A., Beyer, L.: S4l: Self-supervised semi-supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1476–1485 (2019)

[57] Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing **26**(7), 3142–3155 (2017)

[58] Zhang, K., Zuo, W., Zhang, L.: Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. IEEE Transactions on Image Processing **27**(9), 4608–4622 (2018)

[59] Tian, C., Xu, Y., Fei, L., Wang, J., Wen, J., Luo, N.: Enhanced cnn for image denoising. CAAI Transactions on Intelligence Technology **4**(1), 17–23 (2019)

[60] Zhao, H., Shao, W., Bao, B., Li, H.: A simple and robust deep convolutional approach to blind image denoising. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 0–0 (2019)

[61] Tian, C., Xu, Y., Zuo, W.: Image denoising using deep cnn with batch renormalization. Neural Networks **121**, 461–473 (2020)

[62] Tian, C., Xu, Y., Li, Z., Zuo, W., Fei, L., Liu, H.: Attention-guided cnn for image denoising. Neural Networks **124**, 117–129 (2020)

[63] Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2noise: Learning image restoration without clean data. arXiv preprint arXiv:1803.04189 (2018)

[64] Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9446–9454 (2018)

[65] Krull, A., Buchholz, T.-O., Jug, F.: Noise2void-learning denoising from

single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2129–2137 (2019)

[66] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention, pp. 234–241 (2015). Springer

[67] Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., *et al.*: Content-aware image restoration: pushing the limits of fluorescence microscopy. Nature methods **15**(12), 1090–1097 (2018)

[68] Batson, J., Royer, L.: Noise2self: Blind denoising by self-supervision. In: International Conference on Machine Learning, pp. 524–533 (2019). PMLR

[69] Laine, S., Karras, T., Lehtinen, J., Aila, T.: High-quality self-supervised deep image denoising. Advances in Neural Information Processing Systems **32** (2019)

[70] Krull, A., Vičar, T., Prakash, M., Lalit, M., Jug, F.: Probabilistic noise2void: Unsupervised content-aware denoising. Frontiers in Computer Science **2**, 5 (2020)

[71] Prakash, M., Lalit, M., Tomancak, P., Krul, A., Jug, F.: Fully unsupervised probabilistic noise2void. In: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 154–158 (2020). IEEE

[72] Goncharova, A.S., Honigmann, A., Jug, F., Krull, A.: Improving blind spot denoising for microscopy. In: European Conference on Computer Vision, pp. 380–393 (2020). Springer

[73] Moran, N., Schmidt, D., Zhong, Y., Coady, P.: Noisier2noise: Learning to denoise from unpaired noisy data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12064–12072 (2020)

[74] Xu, J., Huang, Y., Cheng, M.-M., Liu, L., Zhu, F., Xu, Z., Shao, L.: Noisy-as-clean: Learning self-supervised denoising from corrupted image. IEEE Transactions on Image Processing **29**, 9316–9329 (2020)

[75] Huang, T., Li, S., Jia, X., Lu, H., Liu, J.: Neighbor2neighbor: Self-supervised denoising from single noisy images. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14781–14790 (2021)

[76] Pang, T., Zheng, H., Quan, Y., Ji, H.: Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2043–2052 (2021)

[77] Quan, Y., Chen, M., Pang, T., Ji, H.: Self2self with dropout: Learning self-supervised denoising from single image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1890–1898 (2020)

[78] Metzler, C.A., Mousavi, A., Heckel, R., Baraniuk, R.G.: Unsupervised learning with stein's unbiased risk estimator. arXiv preprint arXiv:1805.10531 (2018)

[79] Soltanayev, S., Chun, S.Y.: Training deep learning based denoisers without ground truth data. Advances in neural information processing systems **31** (2018)

[80] Lequyer, J., Philip, R., Sharma, A., Pelletier, L.: Noise2fast: Fast self-supervised single image blind denoising. arXiv preprint arXiv:2108.10209 (2021)

[81] Wang, Z., Liu, J., Li, G., Han, H.: Blind2unblind: Self-supervised image denoising with visible blind spots. arXiv preprint arXiv:2203.06967 (2022)

[82] Wang, F., Henninen, T.R., Keller, D., Erni, R.: Noise2atom: unsupervised denoising for scanning transmission electron microscopy images. Applied Microscopy **50**(1), 1–9 (2020)

[83] Desai, A.D., Ozturkler, B.M., Sandino, C.M., Vasanawala, S., Hargreaves, B.A., Re, C.M., Pauly, J.M., Chaudhari, A.S.: Noise2recon: A semi-supervised framework for joint mri reconstruction and denoising. arXiv preprint arXiv:2110.00075 (2021)

[84] Xie, Y., Wang, Z., Ji, S.: Noise2same: Optimizing a self-supervised bound for image denoising. Advances in Neural Information Processing Systems **33**, 20320–20330 (2020)

[85] Hendriksen, A.A., Pelt, D.M., Batenburg, K.J.: Noise2inverse: Self-supervised deep convolutional denoising for tomography. IEEE Transactions on Computational Imaging **6**, 1320–1335 (2020)

[86] Kim, K., Ye, J.C.: Noise2score: Tweedie's approach to self-supervised image denoising without clean images. Advances in Neural Information Processing Systems **34** (2021)

[87] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., Shao, L.: Learning enriched features for real image restoration and enhancement. In: European Conference on Computer Vision, pp. 492–511 (2020). Springer

[88] Tian, C., Xu, Y., Zuo, W., Du, B., Lin, C.-W., Zhang, D.: Designing and training of a dual cnn for image denoising. Knowledge-Based Systems **226**, 106949 (2021)

[89] Mataev, G., Milanfar, P., Elad, M.: Deepred: Deep image prior powered by red. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, pp. 0–0 (2019)

[90] Zheng, H., Yong, H., Zhang, L.: Deep convolutional dictionary learning for image denoising. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 630–641 (2021)

[91] Cheng, S., Wang, Y., Huang, H., Liu, D., Fan, H., Liu, S.: Nbnet: Noise basis learning for image denoising with subspace projection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4896–4906 (2021)

[92] Jang, G., Lee, W., Son, S., Lee, K.M.: C2n: Practical generative noise modeling for real-world denoising. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2350–2359 (2021)

[93] Zuo, Z., Chen, X., Xu, H., Li, J., Liao, W., Yang, Z.-X., Wang, S.: Ideanet: Adaptive dual self-attention network for single image denoising. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 739–748 (2022)

[94] Lyu, Z., Chen, Y., Hou, Y., Zhang, C.: Nstbnet: Toward a nonsubsampled shearlet transform for broad convolutional neural network image denoising. Digital Signal Processing, 103407 (2022)

[95] Xu, M., Xie, X.: Nfcnn: toward a noise fusion convolutional neural network for image denoising. Signal, Image and Video Processing **16**(1), 175–183 (2022)

[96] Fan, L., Li, H., Shi, M., Hua, Z., Zhang, C.: Two-stage image denoising via an enhanced low-rank prior. Journal of Scientific Computing **90**(1), 1–31 (2022)

[97] Jiang, Y., Wronski, B., Mildenhall, B., Barron, J., Wang, Z., Xue, T.: Fast and high-quality image denoising via malleable convolutions. arXiv preprint arXiv:2201.00392 (2022)

[98] Neshatavar, R., Yavartanoo, M., Son, S., Lee, K.M.: Cvf-sid: Cyclic multi-variate function for self-supervised image denoising by disentangling noise from image. arXiv preprint arXiv:2203.13009 (2022)

[99] Ghahremani, M., Khateri, M., Sierra, A., Tohka, J.: Adversarial distortion learning for medical image denoising. arXiv preprint arXiv:2204.14100 (2022)

[100] Wang, Z., Cun, X., Bao, J., Liu, J.: Uformer: A general u-shaped transformer for image restoration. arXiv preprint arXiv:2106.03106 (2021)

[101] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[102] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)

[103] Yang, F., Yang, H., Fu, J., Lu, H., Guo, B.: Learning texture transformer network for image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5791–5800 (2020)

[104] Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12299–12310 (2021)

[105] Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1833–1844 (2021)

[106] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B.: Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 10012–10022 (2021)

[107] Luthra, A., Sulakhe, H., Mittal, T., Iyer, A., Yadav, S.: Eformer: Edge enhancement based transformer for medical image denoising. arXiv preprint arXiv:2109.08044 (2021)

[108] Wang, D., Wu, Z., Yu, H.: Ted-net: Convolution-free t2t vision transformer-based encoder-decoder dilation network for low-dose ct

denoising. In: International Workshop on Machine Learning in Medical Imaging, pp. 416–425 (2021). Springer

[109] Putra, B.P.E.: Salt and pepper noise image denoising dengan menggunakan metode convolutional vision transformer (cvt) (2022)

[110] Prayuda, A.W.H., Prasetyo, H., Guo, J.-M.: Awgn-based image denoiser using convolutional vision transformer. In: 2021 International Symposium on Electronics and Smart Devices (ISESD), pp. 1–6 (2021). IEEE

[111] Fan, C.-M., Liu, T.-J., Liu, K.-H.: Sunet: Swin transformer unet for image denoising. arXiv preprint arXiv:2202.14009 (2022)

[112] Zhang, L., Xiao, Z., Zhou, C., Yuan, J., He, Q., Yang, Y., Liu, X., Liang, D., Zheng, H., Fan, W., *et al.*: Spatial adaptive and transformer fusion network (stfnet) for low-count pet blind denoising with mri. Medical Physics **49**(1), 343–356 (2022)

[113] Yao, C., Jin, S., Liu, M., Ban, X.: Dense residual transformer for image denoising. Electronics **11**(3), 418 (2022)

[114] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H.: Restormer: Efficient transformer for high-resolution image restoration. In: CVPR (2022)

[115] Wang, D., Fan, F., Wu, Z., Liu, R., Wang, F., Yu, H.: Ctformer: Convolution-free token2token dilated vision transformer for low-dose ct denoising. arXiv preprint arXiv:2202.13517 (2022)

[116] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Advances in neural information processing systems **27** (2014)

[117] Tripathi, S., Lipton, Z.C., Nguyen, T.Q.: Correction by projection: Denoising images with generative adversarial networks. arXiv preprint arXiv:1803.04477 (2018)

[118] Lipton, Z.C., Tripathi, S.: Precise recovery of latent vectors from generative adversarial networks. arXiv preprint arXiv:1702.04782 (2017)

[119] Creswell, A., Bharath, A.A.: Inverting the generator of a generative adversarial network. IEEE transactions on neural networks and learning systems **30**(7), 1967–1974 (2018)

[120] Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint

arXiv:1511.06434 (2015)

[121] Kim, D.-W., Ryun Chung, J., Jung, S.-W.: Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0 (2019)

[122] Tran, L.D., Nguyen, S.M., Arai, M.: Gan-based noise model for denoising real images. In: Proceedings of the Asian Conference on Computer Vision (2020)

[123] Wang, F., Xu, Z., Ni, W., Chen, J., Pan, Z.: An adaptive learning image denoising algorithm based on eigenvalue extraction and the gan model. Computational Intelligence and Neuroscience **2022** (2022)

[124] Park, H.S., Baek, J., You, S.K., Choi, J.K., Seo, J.K.: Unpaired image denoising using a generative adversarial network in x-ray ct. IEEE Access **7**, 110414–110425 (2019)

[125] Chen, J., Chen, J., Chao, H., Yang, M.: Image blind denoising with generative adversarial network based noise modeling. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3155–3164 (2018)

[126] Li, Z., Zhou, S., Huang, J., Yu, L., Jin, M.: Investigation of low-dose ct image denoising using unpaired deep learning methods. IEEE Transactions on Radiation and Plasma Medical Sciences **5**(2), 224–234 (2020)

[127] Dong, Z., Liu, G., Ni, G., Jerwick, J., Duan, L., Zhou, C.: Optical coherence tomography image denoising using a generative adversarial network with speckle modulation. Journal of biophotonics **13**(4), 201960135 (2020)

[128] Zhang, S., Wang, L., Chang, C., Liu, C., Zhang, L., Cui, H.: An image denoising method based on bm4d and gan in 3d shearlet domain. Mathematical Problems in Engineering **2020** (2020)

[129] Sengupta, S., Singh, A., Lakshminarayanan, V.: Edgewavenet: edge aware residual wavelet gan for oct image denoising. In: Medical Imaging 2021: Imaging Informatics for Healthcare, Research, and Applications, vol. 11601, p. 116010 (2021). International Society for Optics and Photonics

[130] Vo, D.M., Nguyen, D.M., Le, T.P., Lee, S.-W.: Hi-gan: A hierarchical generative adversarial network for blind denoising of real photographs. Information Sciences **570**, 225–240 (2021)

[131] Wang, Z., Wang, L., Duan, S., Li, Y.: An image denoising method based on deep residual gan. In: Journal of Physics: Conference Series, vol. 1550, p. 032127 (2020). IOP Publishing

[132] Dey, R., Bhattacharjee, D., Nasipuri, M.: Image denoising using generative adversarial network. In: Intelligent Computing: Image Processing Based Applications, pp. 73–90. Springer, ??? (2020)

[133] Fu, Z., Yu, X., Ge, C., Aziz, M.Z., Liu, L.: Adgan: An asymmetric despeckling generative adversarial network for unpaired oct image speckle noise reduction. In: 2021 IEEE 6th Optoelectronics Global Conference (OGC), pp. 212–216 (2021). IEEE

[134] Nagano, Y., Kikuta, Y.: Srgan for super-resolving low-resolution food images. In: Proceedings of the Joint Workshop on Multimedia for Cooking and Eating Activities and Multimedia Assisted Dietary Management, pp. 33–37 (2018)

[135] Bulat, A., Yang, J., Tzimiropoulos, G.: To learn image super-resolution, use a gan to learn how to do image degradation first. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 185–200 (2018)

[136] Zhu, X., Zhang, L., Zhang, L., Liu, X., Shen, Y., Zhao, S.: Gan-based image super-resolution with a novel quality loss. Mathematical Problems in Engineering **2020** (2020)

[137] Zhu, J., Yang, G., Lio, P.: How can we make gan perform better in single medical image super-resolution? a lesion focused multi-scale approach. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), pp. 1669–1673 (2019). IEEE

[138] Dharejo, F.A., Deeba, F., Zhou, Y., Das, B., Jatoi, M.A., Zawish, M., Du, Y., Wang, X.: Twist-gan: Towards wavelet transform and transferred gan for spatio-temporal single image super resolution. ACM Transactions on Intelligent Systems and Technology (TIST) **12**(6), 1–20 (2021)

[139] Jiang, M., Zhi, M., Wei, L., Yang, X., Zhang, J., Li, Y., Wang, P., Huang, J., Yang, G.: Fa-gan: Fused attentive generative adversarial networks for mri image super-resolution. Computerized Medical Imaging and Graphics **92**, 101969 (2021)

[140] de Farias, E.C., Di Noia, C., Han, C., Sala, E., Castelli, M., Rundo, L.: Impact of gan-based lesion-focused medical image super-resolution on the robustness of radiomic features. Scientific reports **11**(1), 1–12 (2021)

[141] Bell-Kligler, S., Shocher, A., Irani, M.: Blind super-resolution kernel estimation using an internal-gan. Advances in Neural Information Processing Systems **32** (2019)

[142] Wang, Y., Perazzi, F., McWilliams, B., Sorkine-Hornung, A., Sorkine-Hornung, O., Schroers, C.: A fully progressive approach to single-image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 864–873 (2018)

[143] Liu, S., Yang, Y., Li, Q., Feng, H., Xu, Z., Chen, Y., Liu, L.: Infrared image super resolution using gan with infrared image prior. In: 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), pp. 1004–1009 (2019). IEEE

[144] Park, S.-J., Son, H., Cho, S., Hong, K.-S., Lee, S.: Srfeat: Single image super-resolution with feature discrimination. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 439–455 (2018)

[145] Gu, H., Unarta, I.C., Huang, X., Yao, Y.: Robust autoencoder gan for cryo-em image denoising. arXiv preprint arXiv:2008.07307 (2020)

[146] Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., Ganguli, S.: Deep unsupervised learning using nonequilibrium thermodynamics. In: International Conference on Machine Learning, pp. 2256–2265 (2015). PMLR

[147] Song, Y., Ermon, S.: Generative modeling by estimating gradients of the data distribution. Advances in neural information processing systems **32** (2019)

[148] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. Advances in neural information processing systems **33**, 6840–6851 (2020)

[149] Jarzynski, C.: Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach. Physical Review E **56**(5), 5018 (1997)

[150] Neal, R.M.: Annealed importance sampling. Statistics and computing **11**, 125–139 (2001)

[151] Nichol, A.Q., Dhariwal, P.: Improved denoising diffusion probabilistic models. In: International Conference on Machine Learning, pp. 8162–8171 (2021). PMLR

[152] Choi, J., Kim, S., Jeong, Y., Gwon, Y., Yoon, S.: Ilvr: Conditioning method for denoising diffusion probabilistic models. arXiv preprint arXiv:2108.02938 (2021)

[153] Lugmayr, A., Danelljan, M., Romero, A., Yu, F., Timofte, R., Van Gool, L.: Repaint: Inpainting using denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11461–11471 (2022)

[154] Song, J., Meng, C., Ermon, S.: Denoising diffusion implicit models. arXiv preprint arXiv:2010.02502 (2020)

[155] Sasaki, H., Willcocks, C.G., Breckon, T.P.: Unit-ddpm: Unpaired image translation with denoising diffusion probabilistic models. arXiv preprint arXiv:2104.05358 (2021)

[156] Gong, K., Johnson, K.A., Fakhri, G.E., Li, Q., Pan, T.: Pet image denoising based on denoising diffusion probabilistic models. arXiv preprint arXiv:2209.06167 (2022)

[157] Dorjsembe, Z., Odonchimed, S., Xiao, F.: Three-dimensional medical image synthesis with denoising diffusion probabilistic models. In: Medical Imaging with Deep Learning (2022)

[158] Khader, F., Mueller-Franzes, G., Arasteh, S.T., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baessler, B., Foersch, S., et al.: Medical diffusion–denoising diffusion probabilistic models for 3d medical image generation. arXiv preprint arXiv:2211.03364 (2022)

[159] Nair, N.G., Mei, K., Patel, V.M.: At-ddpm: Restoring faces degraded by atmospheric turbulence using denoising diffusion probabilistic models. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3434–3443 (2023)

[160] Khader, F., Müller-Franzes, G., Tayebi Arasteh, S., Han, T., Haarburger, C., Schulze-Hagen, M., Schad, P., Engelhardt, S., Baeßler, B., Foersch, S., *et al.*: Denoising diffusion probabilistic models for 3d medical image generation. Scientific Reports **13**(1), 7303 (2023)

[161] Xie, Y., Li, Q.: Measurement-conditioned denoising diffusion probabilistic model for under-sampled medical image reconstruction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 655–664 (2022). Springer

[162] Sahak, H., Watson, D., Saharia, C., Fleet, D.: Denoising diffusion probabilistic models for robust image super-resolution in the wild. arXiv preprint arXiv:2302.07864 (2023)

[163] Pan, S., Wang, T., Qiu, R.L., Axente, M., Chang, C.-W., Peng, J., Patel, A.B., Shelton, J., Patel, S.A., Roper, J., *et al.*: 2d medical image synthesis using transformer-based denoising diffusion probabilistic model. Physics

in Medicine & Biology **68**(10), 105004 (2023)

[164] Vo, D.M., Le, T.P., Nguyen, D.M., Lee, S.-W.: Boostnet: A boosted convolutional neural network for image blind denoising. IEEE Access **9**, 115145–115164 (2021)

[165] Zhang, K., Li, Y., Liang, J., Cao, J., Zhang, Y., Tang, H., Fan, D.-P., Timofte, R., Gool, L.V.: Practical blind image denoising via swin-conv-unet and data synthesis. Machine Intelligence Research (2023). https://doi.org/10.1007/s11633-023-1466-0

[166] Rahman, S.S.M.M., Salomon, M., DembÉlÉ, S.: Noise analysis to guide denoising of scanning electron microscopy images. In: 2023 9th International Conference on Control, Decision and Information Technologies (CoDIT), pp. 1559–1564 (2023). IEEE

[167] Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep cnn denoiser prior for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3929–3938 (2017)

[168] Hui, Z., Gao, X., Yang, Y., Wang, X.: Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th Acm International Conference on Multimedia, pp. 2024–2032 (2019)

[169] Mehri, A., Ardakani, P.B., Sappa, A.D.: Mprnet: Multi-path residual network for lightweight image super resolution. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 2704–2713 (2021)

[170] Zhang, K., Liang, J., Van Gool, L., Timofte, R.: Designing a practical degradation model for deep blind image super-resolution. In: IEEE International Conference on Computer Vision, pp. 4791–4800 (2021)

[171] Zhang, Y., Li, D., Law, K.L., Wang, X., Qin, H., Li, H.: Idr: Self-supervised image denoising via iterative data refinement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 2098–2107 (2022)

[172] Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. arXiv preprint arXiv:2204.04676 (2022)

[173] Zamir, S.W., Arora, A., Khan, S., Hayat, M., Khan, F.S., Yang, M.-H., Shao, L.: Learning enriched features for fast image restoration and enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2022)

[174] Yang, C., Liang, L., Su, Z.: Real-world denoising via diffusion model.

arXiv preprint arXiv:2305.04457 (2023)

[175] Aversa, R., Modarres, M.H., Cozzini, S., Ciancio, R., Chiusole, A.: The first annotated set of scanning electron microscopy images for nanoscience. Scientific data **5**(1), 1–10 (2018)

[176] Wang, X., Xie, L., Dong, C., Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1905–1914 (2021)

[177] Zhang, K., Van Gool, L., Timofte, R.: Deep unfolding network for image super-resolution. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3217–3226 (2020)

[178] Rahman, S.S.M.M., Salomon, M., Dembélé, S.: Machine learning aided classification of noise distribution in scanning electron microscopy images. In: 2023 3rd International Conference on Computer, Control and Robotics (ICCCR), pp. 111–115 (2023). IEEE

[179] Alcantarilla, P., Nuevo, J., Bartoli, A.: Fast explicit diffusion for accelerated features in nonlinear scale spaces british machine vision conference (BMVC). Bristol (2013)

[180] Dong, G., Basu, A.: Medical image denosing via explainable ai feature preserving loss. arXiv preprint arXiv:2310.20101 (2023)

[181] Zhang, K.: GitHub - cszn/KAIR: Image Restoration Toolbox (PyTorch). Training and testing codes for DPIR, USRNet, DnCNN, FFDNet, SRMD, DPSR, BSRGAN, SwinIR — github.com. https://github.com/cszn/KAIR. [Accessed 21-01-2024]