

I would like to thank my supervisor, Dr. Carlo Lucibello, for providing valuable guidance and feedback throughout this project.

I would also like to thank Dr. Filippo Molinari and Dr. Kristen M. Meiburger from the Politecnico di Torino who kindly shared the CUBS dataset and their knowledge in the field of medical imaging with me.

Contents

1	Introduction	2
1.1	Diffusion probabilistic models	3
1.2	DDPMs original formulation	3
2	Diffusion model formulation for image denoising	7
2.1	Unsupervised denoising methods	8
2.1.1	A review of current unsupervised denoising methods	9
2.1.2	Unsupervised denoising with diffusion models	10
2.1.3	Example: Diffusion BM3D	11
2.1.4	Diffusion BM3D performance on BSD68 Dataset	12
2.1.5	Results comparison	15
2.2	Supervised denoising methods	16
2.2.1	Current supervised denoising algorithms	17
2.2.2	Diffusion formulation for supervised denoising	18
2.2.3	Example: Diffusion U-Net	19
2.2.4	Generalization	19
3	Applications in medical imaging	21
3.1	Issues with medical imaging	21
3.1.1	Noise approximation	22
3.2	CUBS dataset: Carotid Ultrasound Boundary Study Dataset	23
3.2.1	Unsupervised denoising methods results on CUBS dataset	24
3.2.2	Finding the clean distribution: diffusion BM3D	25
3.2.3	Learning the clean distribution: diffusion U-Net	26
3.2.4	Medical dataset denoising pipeline	27
4	Conclusions	29

1 Introduction

Diffusion models are a class of generative models that have recently been used in various applications, including image and video generation, text-to-image synthesis, and more [1, 2, 3, 4, 5]. They have also shown promising results in medical imaging applications such as MRI reconstruction and image super-resolution[6, 7]. In this thesis work, I analyse their use in the application of denoising, particularly in the context of medical imaging denoising [8], to enhance the quality and accuracy of medical diagnostics.

Images captured by different medical devices contain intrinsic artefacts that can limit the diagnostic performance of these techniques [9]. Ultrasound, CT and MRI scans often contain speckle noise. Speckle noise is a granular interference inherent in most imaging systems caused by wave interference during tissue interaction. This noise, characterized by its granular appearance, degrades medical image quality by reducing contrast and obscuring fine details, thus complicating diagnosis[10]. An optimal de-speckling algorithm should remove speckle noise in homogeneous areas of the image, while the edges in the image should be preserved [11]. To achieve this result, it can be shown that a diffusion model formulation can improve the performance of denoising algorithms and better preserve the finer details of the images by iterating over different noise levels during its sampling procedure.

The work is organized into two main parts. The first part introduces the fundamental concept of diffusion models and their formulation in the denoising problem. Next I show how diffusion can be applied to unsupervised and supervised denoising and propose a variation of the algorithm BM3D [12] that achieves state-of-the-art results using the diffusion sampling formulation. I discuss the application of unsupervised and supervised diffusion-based denoising methods to the problem of medical image denoising where it is difficult to obtain high-resolution images for use as clean images, and propose a pipeline using the CUBS, Carotid Ultrasound Boundary Study, dataset [13, 14]. Finally, I draw some conclusions by summarizing the main results obtained.

1.1 Diffusion probabilistic models

A diffusion probabilistic model is a parameterized Markov chain trained using variational inference to produce samples that match the data distribution after a finite number of steps [1, 15]. This model consists of two primary components: a forward pass, progressively adding noise to the data, and a backward pass, that attempts to reconstruct the original distribution from the noisy one. The backward pass is also used as a sampling procedure to create new clean data in generative models. The specific formulation of the forward and backwards pass can vary depending on the model’s design and intended use. Below is the original proposed formulation for one of the earliest diffusion models and the one still most commonly used for image generation. In section 2, I propose how to modify the structure for the real image denoising task.

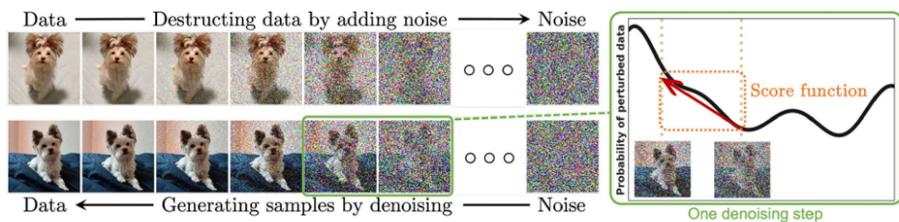


Figure 1: Forward and backwards passes in Generative Diffusion Models. The reverse step requires a score function to optimally denoise the data (seen in the figure on the right) [16]

1.2 DDPMs original formulation

In 2020 Ho et al. propose DDPMs, Denoising Diffusion Probabilistic Models, a diffusion model formulation of the forward and backward passes that they successfully apply for image generation [1].

It is worth noting that in DDPMs the term “denoising” comes from the denoising variational autoencoder used in the backwards pass for image generation from the noisy distribution and not from the fact that the model is used for real world image denoising tasks.

Forward pass

The forward pass in DDPMs involves gradually corrupting the data from its original form into pure Gaussian noise $\mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$ through a Markov chain of Gaussian transitions. This process is parameterized by a fixed variance schedule β_t , which dictates the amount of noise added at each step. The process can be mathematically expressed as follows:

$$q(\mathbf{x}_{1:T} | \mathbf{x}_0) := \prod_{t=1}^T q(\mathbf{x}_t | \mathbf{x}_{t-1}), \quad (1)$$

$$q(\mathbf{x}_t | \mathbf{x}_{t-1}) := \mathcal{N}(\mathbf{x}_t; \sqrt{1 - \beta_t} \mathbf{x}_{t-1}, \beta_t \mathbf{I}), \quad (2)$$

The image at time t can thus be obtained from a combination of \mathbf{x}_{t-1} and the Gaussian noise \mathbf{z} with the following formula:

$$\mathbf{x}_t = \sqrt{1 - \beta_t} \mathbf{x}_{t-1} + \sqrt{\beta_t} \mathbf{z}, \quad \mathbf{z} \sim \mathcal{N}(0, \mathbf{I}) \quad (3)$$

Where \mathbf{x}_t is the image at time step t , β_t is the predefined variance of the noise added at step t and \mathbf{z} represents the pure Gaussian noise.

Taking α_t and $\bar{\alpha}_t$ as follows: $\alpha_t := 1 - \beta_t$, $\bar{\alpha}_t := \prod_{s=1}^t \alpha_s$ and starting from \mathbf{x}_0 the forward pass can be rewritten as:

$$q(\mathbf{x}_t | \mathbf{x}_0) = \mathcal{N}(\mathbf{x}_t; \sqrt{\bar{\alpha}_t} \mathbf{x}_0, (1 - \bar{\alpha}_t) \mathbf{I}) \quad (4)$$

Meaning the data \mathbf{x}_t can be obtained by a combination of the starting image \mathbf{x}_0 and the Gaussian noise:

$$\mathbf{x}_t = \sqrt{\bar{\alpha}_t} \mathbf{x}_0 + (1 - \bar{\alpha}_t) \mathbf{z} \quad (5)$$

Backwards pass

The backward pass, or reverse process, is where DDPMs exhibit their generative capabilities. Starting from Gaussian noise, the model aims to reconstruct the original data by reversing the noise addition process described in the forward pass. The backward

pass is modelled using a parameterized neural network that predicts the noise epsilon added at each step of the forward pass, which is then used for iterative denoising the data.

Starting from the final step distribution $p(\mathbf{x}_T) = \mathcal{N}(\mathbf{x}_T; 0, \mathbf{I})$ which is the Gaussian noise distribution, we can obtain the conditional probability distribution of any step $t - 1$ given the distribution at t :

$$p_\theta(\mathbf{x}_{0:T}) := p(\mathbf{x}_T) \prod_{t=1}^T p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t), \quad p_\theta(\mathbf{x}_{t-1}|\mathbf{x}_t) := \mathcal{N}(\mathbf{x}_{t-1}; \mu_\theta(\mathbf{x}_t, t), \Sigma_\theta(\mathbf{x}_t, t)) \quad (6)$$

Where μ_θ is the expected value of the prediction \mathbf{x}_t and Σ_θ is the covariance of the prediction.

By backwards iteration the image \mathbf{x}_{t-1} can be obtained by subtracting the noise prediction of the neural network $\epsilon_\theta(\mathbf{x}_t, t)$ from the image \mathbf{x}_t .

$$\mathbf{x}_{t-1} = \frac{1}{\sqrt{1 - \beta_t}} \left(\mathbf{x}_t - \sqrt{\beta_t} \epsilon_\theta(\mathbf{x}_t, t) \right) \quad (7)$$

Where $\epsilon_\theta(\mathbf{x}_t, t)$ is the noise predicted by the neural network, parameterized by θ and the term $\sqrt{\beta_t}$ scales the predicted noise to the appropriate level based on the forward process's fixed variance schedule.

Training

The training objective in DDPMs focuses on optimizing the neural network $\epsilon_\theta(\mathbf{x}_t, t)$ to accurately predict the noise at each step added during the forward pass defined by ϵ_t . This is typically done by minimizing the mean squared error between the noise added during the forward process and the noise predicted by the neural network during the backward process:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0, t, \epsilon} [\|\epsilon_t - \epsilon_\theta(\mathbf{x}_t, t)\|^2] \quad (8)$$

Sampling

Once the neural network is trained, new clean samples can be obtained that, ideally, have the same distribution as the clean images in the training set. To do so, a new image $\tilde{\mathbf{x}}_T$ is sampled from the Gaussian noise distribution and the backwards pass formulation is applied to the image iteratively, removing the noise predicted by the neural network at each step.

At time zero, all the predicted noise has been removed by the network and the image $\tilde{\mathbf{x}}_0$ is a new generated image with the same distribution as the clean image. Since the Gaussian noise at the start was randomly generated, this new image does not exactly match any of the clean images in the training set but presents the same features.

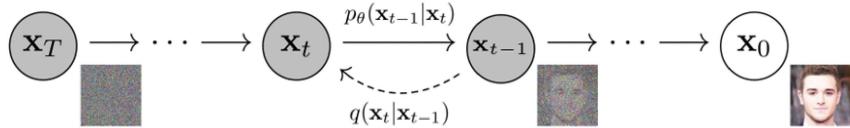


Figure 2: Backward and forward procedure formulation in DDPMs[1]

2 Diffusion model formulation for image denoising

In this section, I illustrate a modification to the DDPMs forward and backward passes to adapt the diffusion model formulation to the image denoising task, starting from the concepts proposed by Yang et al. in 2023 [17]. In their paper, Yang et al. propose a modification to the standard forward and backward passes to adapt the diffusion model formulation to the image denoising task. In their formulation, the final step of the forward pass and the first step of the backward pass are no longer an image with Gaussian noise, but rather the noisy image. The noisy image consists of the clean image component plus an added noise image that is assumed to be Gaussian noise.

In this formulation, \mathbf{x}_0 represents the clean image and $\mathbf{y} = \mathbf{x}_0 + \mathbf{z}$ represents the noisy image, where $\mathbf{z} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ is the added Gaussian noise with zero mean and a defined standard deviation σ .

Forward pass

The forward pass progressively adds noise, interpolating between the clean image \mathbf{x}_0 and the noisy image \mathbf{y} , effectively adding fractions of the Gaussian noise \mathbf{z} until the noisy image \mathbf{y} is obtained. Formally, for $t = 0, \dots, T$:

$$\mathbf{x}_t = (1 - \alpha_t)\mathbf{x}_0 + \alpha_t\mathbf{y} \quad \text{with } \alpha_t = \left(\frac{t}{T}\right)^{\exp} \quad (9)$$

Here, \mathbf{x}_t represents the image at time step t of the forward pass. The standard deviation of its noise component is $\alpha_t\sigma$. The noise schedule α_t controls the interpolation between the clean image \mathbf{x}_0 and the noisy image \mathbf{x}_t . If the exponent is 1, the interpolation is linear.

Backward pass

The backward pass involves training a neural network or using another algorithm to predict the clean image distribution from the noisy image at a specific time step t . The pass starts at $t = T$ with $\mathbf{x}'_T = \mathbf{y}$. Then, for $t = T, \dots, 0$:

$$\mathbf{x}'_{t-1} = (1 - \alpha_{t-1})S(\mathbf{x}'_t, t) + \alpha_{t-1}\mathbf{y} \quad \text{with } \alpha_{t-1} = \left(\frac{t-1}{T}\right)^{\exp} \quad (10)$$

Where \mathbf{x}'_t represents the denoised image at time step t . S denotes a denoising algorithm that takes as input the noisy image \mathbf{x}'_t and the time step t and outputs a predicted clean image. Here, α_{t-1} controls the interpolation between the network's output $S(\mathbf{x}_t, t)$ and the noisy image \mathbf{y} .

Training objective

When using a neural network in the backwards pass, the objective in the training procedure is to accurately restore the clean image \mathbf{x}_0 . This is achieved by minimizing a loss function that measures the discrepancy between the restored image and the original clean image at every time step t , e.g. the mean squared error. In this case the loss function can be written as:

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}_0, \mathbf{y}, t} [\|\mathbf{x}_0 - S(\mathbf{x}_t, t)\|^2] \quad (11)$$

where θ are the parameters of the neural network S .

Sampling procedure

The sampling procedure involves progressively moving from a noisy image to a clean image by using the trained neural network at different time steps. Starting from a noisy image \mathbf{x}_T that exhibits the same noise pattern as the training set, it is possible to iteratively apply the backward pass formula to generate cleaner versions of the image until reaching \mathbf{x}_0 .

2.1 Unsupervised denoising methods

Unsupervised denoising algorithms are specifically designed to remove the noise from an image without any clean sample to learn from [18]. They are particularly useful in cases where it is very difficult or even impossible to obtain accurate clean images. This is often the case with medical imaging and is the reason why historically unsupervised denoising algorithms have been the most common in this field [19].

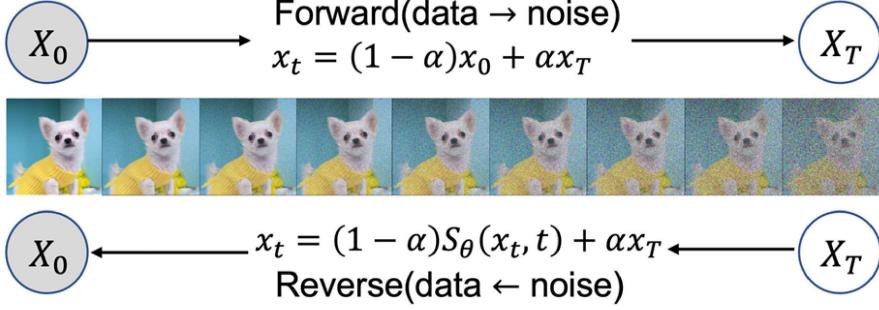


Figure 3: Forward and backward passes in diffusion model formulation for image denoising, the image at X_T is not random Gaussian noise but the noisy image instead[17]

2.1.1 A review of current unsupervised denoising methods

Several unsupervised denoising techniques have been implemented throughout the years. For example, Non-Local Means (NLM) denoises an image by averaging the values of similar patches, capitalizing on the repetitive patterns within the image to distinguish noise from the signal [20]. Total Variation Denoising (TV) reduces noise by minimizing the total variation of the image, which effectively preserves edges while smoothing out noise-induced variations [21]. Wavelet Transform Denoising transforms the image into the wavelet domain, where noise and signal can be separated based on their frequency characteristics, allowing for threshold that suppresses noise while retaining important features [22].

Principal Component Analysis (PCA) reconstructs the image using only the principal components that capture the most variance, assuming that noise corresponds to components with lesser variance [23]. Dictionary Learning and Sparse Coding involve learning a set of basis functions that can sparsely represent image patches, thereby filtering out noise through sparse representation [24]. Block-Matching and 3D Filtering (BM3D) groups similar image patches into 3D arrays and collaboratively filters them using transform-domain shrinkage, combining both spatial and frequency domain processing to achieve state-of-the-art denoising performance [25].

In the latest years, unsupervised neural networks have been proposed for the very same reason. Networks like Noise2Void are capable of learning a noise profile and consequently remove it from the image by learning from the image itself [26]. Autoencoders learn an efficient representation of the data by encoding and then decoding the input

image, effectively removing noise in the process [27].

2.1.2 Unsupervised denoising with diffusion models

In this section, I propose a method to integrate the sampling procedure of diffusion models with unsupervised denoising algorithms. This approach allows for the prediction of denoised images at each step of the backward pass, gradually reducing the standard deviation of the noise profile. The unsupervised denoising algorithm requires the standard deviation of the noise (or a similar metric) as input. The denoising function is denoted as $D(X, \sigma)$, where X is the image to be denoised and σ is its standard deviation.

Starting from the backward pass in our denoising formulation, it is assumed that the noise level of the noisy image X_t at step T is the given noisy image standard deviation σ_T , and the noise level at time step t is $\sigma_t = \sigma_T(\alpha_t)$. Then the denoising function $D(X_t, \sigma_t)$ is applied, where X_t is the image at step t and σ_t is its noise component standard deviation.

$$X_{t-1} = (1 - \alpha_{t-1})D(X_t, \sigma_t) + \alpha_{t-1}X_t \quad (12)$$

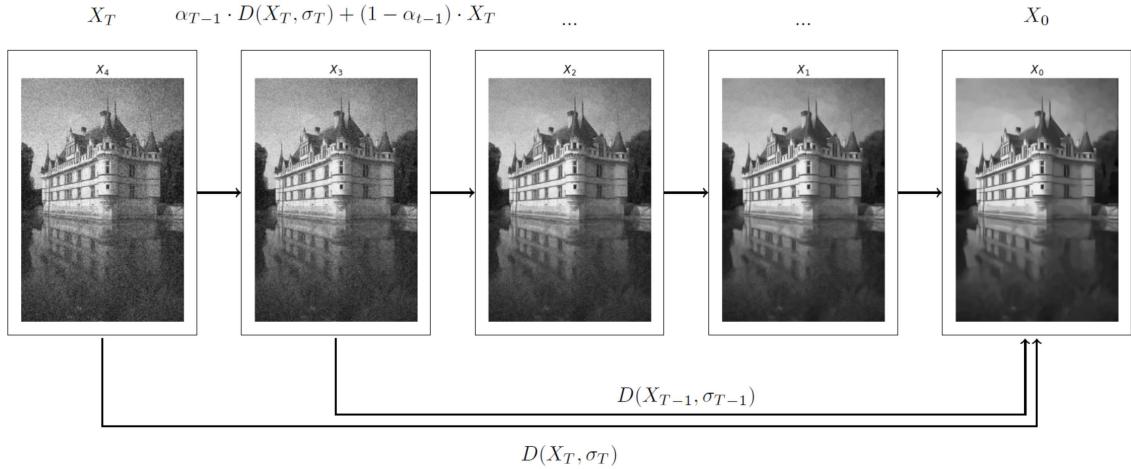


Figure 4: Diffusion sampling procedure for unsupervised denoising using algorithm $D(X_t, \sigma_t)$

2.1.3 Example: Diffusion BM3D

BM3D, Block-Matching and 3D Filtering, is considered to be the state-of-the-art unsupervised denoising algorithm and achieved the best results on image denoising tasks before the introduction of supervised learning methods [12].

BM3D first extracts small overlapping patches from the image, then it groups similar patches by using metrics such as Euclidean distance and stacks the patches into a 3D block. The method then applies a linear transform such as the discrete wavelet transform to each block to separate the noise profile from the signal. Once each block is in transform form, the noise is reduced by using a shrinkage function on the coefficients it identifies as being part of the noise while retaining the ones that it identifies as being significant features of the data. To distinguish between the two, BM3D uses a threshold based on the noise level argument sigma, the higher the noise level the higher the coefficient of a feature in the transform space needs to be not to be identified as noise. Lastly, it uses an inverse 3D transform to convert the denoised 3D block back to the spatial domain. In the second step this process is repeated with Wiener Filtering to smooth out the image [28, 29].

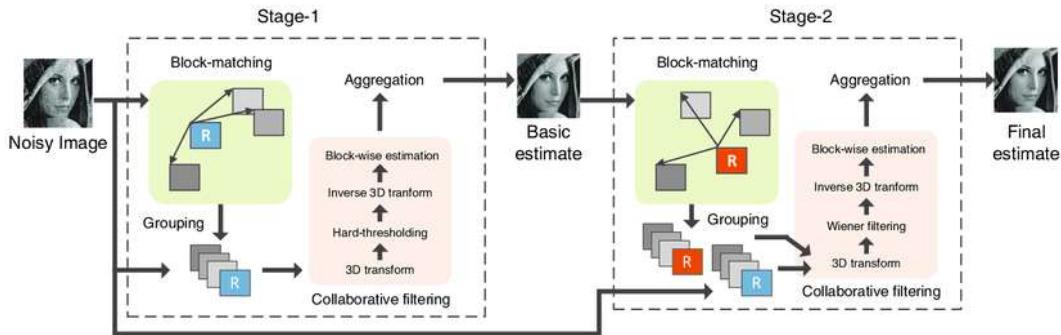


Figure 5: Visual representation of the two steps of the BM3D algorithm[30]

In the following, I introduce a variation of BM3D using the diffusion framework formulation, which I will call “diffusion BM3D”. In this formulation, I apply BM3D with different noise arguments sigma at each time step to progressively denoise the image. By doing this, I do not remove features identified as noise in one step but divide the denoising task into a series of sub-problems, gradually lowering and gradually lower

the noise level and preserving finer details in the image. I use the sampling formulation outlined in Algorithm 1.

Algorithm 1 Diffusion BM3D: denoising procedure

```

1:  $x_T = y$ 
2: for  $t = T, \dots, 1$  do
3:   Compute  $\alpha_{t-1} = \left(\frac{t-1}{T}\right)^{\exp}$ 
4:   Compute  $\sigma_t = \sigma_T \cdot \alpha_t$ 
5:    $x_{t-1} = (1 - \alpha_{t-1})D(x_t, \sigma_t) + \alpha_{t-1}x_t$ 
6: end for
7: return  $x_0$ 
```

2.1.4 Diffusion BM3D performance on BSD68 Dataset

I compute the performance of Diffusion BM3D on the BSD68, Berkeley Segmentation Dataset, which is a benchmark dataset commonly used in the field of image processing and computer vision for evaluating image denoising algorithms, comprising 68 gray-scale images at different noise levels. The most common noise level values used in the dataset for model performance evaluation are 15, 25 and 50. [31]

Performance metrics The metrics commonly used to evaluate denoising algorithms performance are PNSR and SSIM.

PSNR Peak Signal-to-Noise Ratio, is a metric used to assess the quality of a reconstructed image compared to a reference image [32]. It is the most commonly employed metric to measure image denoising performance due to it's ease of computation.

$$\text{PSNR}(x, y) = 20 \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}(x, y)}} \right) \quad (13)$$

Where x is the clean image, y is the reconstructed image, MAX_I is the maximum possible pixel value of the images (255 for 8-bit images), and $\text{MSE}(x, y)$ is the mean squared error between the two images.

SSIM Structural Similarity Index Measure, is a perceptual metric that quantifies image difference in terms of image quality degradation [33]. Differently from PSNR

it considers changes in structural information, luminance, and contrast. Compared to PSNR, SSIM is more aligned with human perception providing a more meaningful interpretation in cases where human interpretation is required. It is not as widely used due to the less straight-forward formulation.

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (14)$$

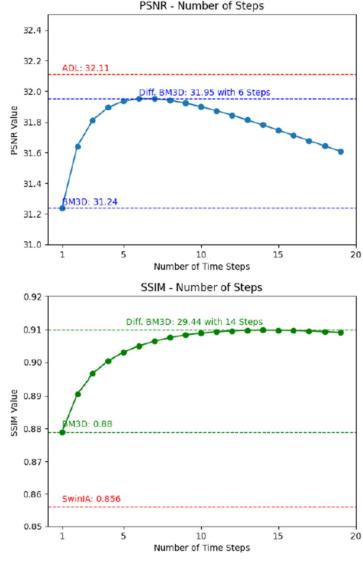
Where μ_x and μ_y are the average of the two images, σ_x^2 and σ_y^2 are their variance, σ_{xy} is the covariance of x and y and C_1, C_2 are small constants to stabilize the division.

Diffusion BM3D with linear schedule: exp=1 First I applied the model with $\text{exp} = 1$, meaning the noise schedule α_t decreases linearly. I then computed its performance on the BSD68 Dataset for every number of steps in the sampling process. In the plots of Figure 6, I report the PSNR and SSIM metrics as the number of steps increases for the three considered noise levels: 15, 25, and 50.

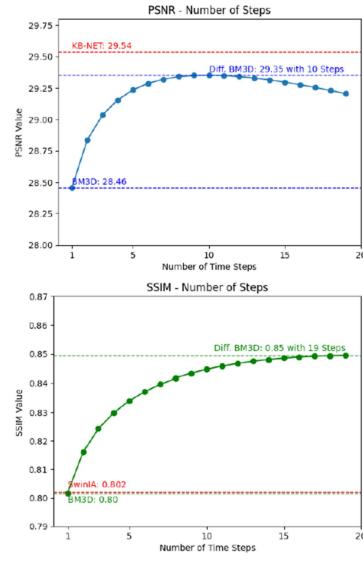
The highest recorded performance by a state-of-the-art model is also reported in red in Figure 6. The data for this is obtained from the "Papers With Code" website[34].

As seen from the plots of Figure 6, applying the diffusion sampling procedure significantly increases the performance of BM3D on the three datasets. The biggest increase in terms of PSNR can be seen in the sigma25 dataset where the PSNR increases from 28.46, with the normal procedure, to 29.35, using the diffusion procedure with 10 steps. However, after a certain number of steps, the performance starts to decline, both in terms of PSNR and SSIM. The optimal number of steps differs based on the starting noise level and target metric. The general trend that can be observed is that, the higher the starting noise, the more steps are needed for the optimal solution. The SSIM-optimal number of steps seems to be higher than the PSNR one as well. The exact reason for this behaviour is not clear but it can be inferred from the trend that when the amount of noise subtracted is below a certain threshold, the model begins to diverge from the optimal denoising pattern. I will go in more details when comparing image results in Section 2.1.5.

BSD68-sigma15



BSD68-sigma25



BSD68-sigma50

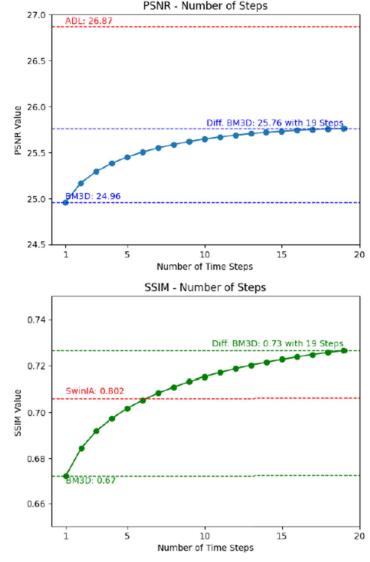


Figure 6: PNSR (top) and SSIM (bottom) values for different number of steps in the Diffusion BM3D procedure (with $\exp=1$) on the three datasets BSD68 sigma15, sigma25, sigma50 (left, middle, right). Values for the baseline BM3D model, optimal number of steps and state-of-the-art model are reported (red).

PNSR and SSIM performance comparison In this paragraph, I compare the performance of the diffusion BM3D model, using the optimal number of steps previously found, with the current unsupervised denoising methods in terms of PNSR and SSIM. I also show that diffusion BM3D comes close with state-of-the-art supervised denoising models in terms of PNSR on the three noise levels of the BSD68 dataset.

Table 1: Comparison of PNSR and SSIM values with unsupervised denoising algorithms on BSD68 datasets [34].

	Sigma 15		Sigma 25		Sigma 50	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Unsupervised Algorithms						
PCA [23]	28.26	0.725	23.94	0.56	18.33	0.331
Wavelet [22]	28.79	0.838	25.54	0.721	22.36	0.547
Non-Local Means [20]	29.81	0.818	26.64	0.731	22.91	0.571
Total Variation Denoising [21]	29.84	0.849	27.12	0.76	25.22	0.696
BM3D [25]	31.44	0.882	28.57	0.801	25.06	0.673
Swin IA (self supervised) [35]	31.07	0.856	29.17	0.802	26.87	0.706
Proposed Method						
Diffusion BM3D	31.95	0.909	29.35	0.849	25.76	0.726

Diffusion BM3D achieves state-of-the-art performance in both PNSR and SSIM on sigma15 and sigma25, improving the values achieved by the second best method SwinIA.

Table 2: Comparison of PSNR values with state-of-the-art denoising methods on BSD68 datasets. SSIM values not reported. [34]

	Sigma 15		Sigma 25		Sigma 50	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
State-of-the-art Methods - Supervised						
ADL [36]	32.11		29.5		26.87	
KB-Net [37]	31.98		29.54		26.65	
MWCNN [38]	31.86		29.41		26.53	
DnCNN [39]	31.63		29.23		26.19	
Proposed Method - Unsupervised						
Diffusion BM3D	31.95		29.35		25.76	

Diffusion BM3D tends to perform worse on the sigma50 dataset, probably due to the lower performance of the baseline model BM3D on the high noise level. When compared to supervised methods, Diffusion BM3D comes close to the PNSR values of state-of-the-art methods on the sigma15 and sigma25 datasets. The benefit of not having to train the model on images with the same noise profile could make Diffusion BM3D a valid alternative to this methods. Unfortunately, SSIM values on the BSD68 dataset are not reported for these models and comparisons are not possible. Out of the reported values, Diffusion BM3D achieves state-of-the-art performance.

Diffusion BM3D with non-linear schedule: $\exp=3/2$ The exponent value in the sampling procedure can also be changed, leading to a non-linear decrease in the noise schedule term α_t . Among the possible values, I found $\exp = \frac{3}{2}$ to be the best performing. This means that the first iterations will have a sharper decrease in the noise level while the last ones will have a slower decline. Figure 7 reports the plots of PSNR and SSIM of Diffusion BM3D with $\exp = \frac{3}{2}$ on the BSD68 dataset with noise level $\sigma = 25$.

In terms of PNSR and SSIM, a non-linear schedule seems to increase the performance of the model in some cases. More importantly, it usually reduces the number of steps required to achieve the optimal solution which reduces computational time. More thorough research is needed to understand the reason behind this and how to choose an optimal noise schedule.

2.1.5 Results comparison

As seen from Figures 6-7, increasing the number of steps in the denoising procedure leads to sharper edges and better quality in the finer details of the image.

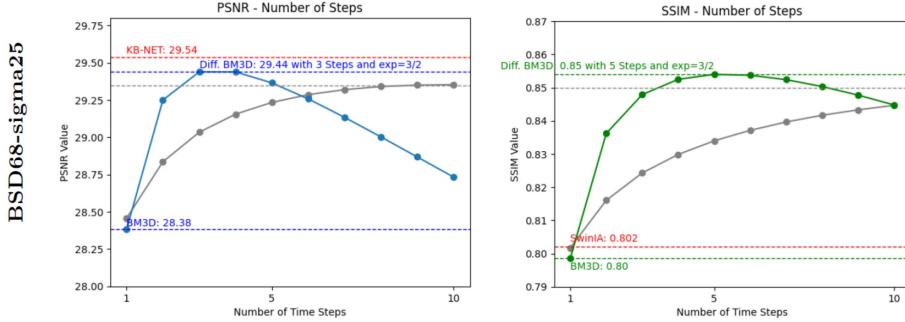


Figure 7: PNSR (left) and SSIM (right) values for different number of steps in the Diffusion BM3D procedure with $\text{exp}=3/2$ on the dataset BSD68 sigma25. The PNSR and SSIM values of the linear Diffusion BM3D for each number of steps are reported in gray.

The drawback is that the method also produces certain artifacts and strong contours that are not present in the original image especially where the edges are soft. This leads to a decline in model performance when the number of steps is very large.



Figure 8: Comparison of resulting images from unsupervised denoising methods and proposed Diff BM3D on the BSD68 sigma25 dataset

2.2 Supervised denoising methods

When training supervised denoising methods, neural networks are first trained on large sets of high-resolution clean images representing all the possible patterns that are likely to appear in camera pictures [40].

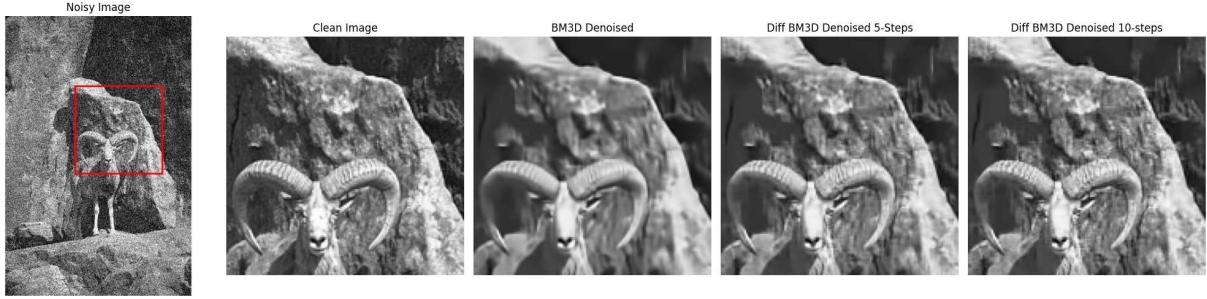


Figure 9: Resulting images from using different number of steps in the Diff BM3D procedure on BSD68 sigma25 (1, 5, 10 steps from left to right)

Artificial noise with the desired profile and standard deviation is then added to these images and the network is trained to remove this noise and learn the original clean pictures. Once trained, the network can be applied to predict an unseen smaller test set with the same standard deviation as the artificial noise added in the training process.

2.2.1 Current supervised denoising algorithms

Current supervised denoising methods leverage deep learning techniques, particularly Convolutional Neural Networks (CNNs) and transformer-based architectures, to learn to remove noisy data and obtain clean distributions [39, 41]. One common approach is deep residual learning, employed by networks like DnCNN (Denoising Convolutional Neural Network) [39]. Another method is the U-Net architecture, which uses a symmetric encoder-decoder structure with skip connections to preserve spatial information and perform important feature extraction, enabling high-quality denoising [42]. GANs, Generative Adversarial Networks, can also be applied to denoising, using a generator network that creates denoised images and a discriminator network that distinguishes between clean and noisy images, leading to realistic noise reduction [43]. Finally, transformer-based models, known for their powerful sequence modelling capabilities, have been adapted for image denoising, offering superior performance by capturing long-range dependencies within the image [41, 44]. The transformer-based model KB3NET uses an advanced multi-scale architecture and a novel attention mechanism to obtain clean samples and achieves state-of-the-art results on most benchmark datasets, outperforming previous methods in terms of PSNR and SSIM values [45].

2.2.2 Diffusion formulation for supervised denoising

This paragraph describes how to apply the diffusion model architecture for training denoising supervised methods. In the paper [17], Yang et al also propose a training procedure for a model called Diffusion U-Net using a formulation like the one described above. This next part aims to better contextualize this process and give a general framework for supervised denoising learning using the formulation. It then contextualizes the models that integrate this procedure and show their performance on different datasets where they achieved state of the art results.

Forward pass: training a time conditioned denoising neural network In the forward pass, iteration is carried out over each pair of clean and noisy images in the training set and noise is added gradually between the clean and noisy images with the formulation of the forward pass. The goal during the training procedure is to train a time conditioned denoising neural network. An existing denoising neural network $S_\theta(x)$ is selected and embed with a time step tensor. For every pair of images and every iteration through the forward pass, the image and the time step at which it was obtained $S_\theta(x_t, t)$ are fed to the network. The neural network will then be trained with gradient descent on a loss function between the denoised image and the clean image. By doing so the network learns the right amount of denoising to apply at any time step and will be trained to denoise different noise level. In the backwards pass, it becomes therefore possible to apply a course-to-fine denoising procedure similar to that of the unsupervised algorithm with this trained network.

Algorithm 2 Training

```
1: repeat
2:    $x_0, x_T \sim$  training set (clean, noisy)
3:   for  $t \in 0, \dots, T$  do
4:     Compute  $\alpha_t = \left(\frac{t}{T}\right)^{\exp}$ 
5:      $x_t = (1 - \alpha_t)x_0 + \alpha_t x_T$ 
6:     Take gradient descent step on  $\nabla_\theta \|x_0 - S_\theta(x_t, t)\|^2$ 
7:   end for
8: until converged
```

Backwards pass: sampling procedure The backwards pass is very similar to the sampling procedure of the unsupervised method. The difference is that it is not necessary to feed the network the specific noise level, only the specific time step of the sampling procedure. To obtain the clean version of an image from unseen data it is sufficient to feed the network a noisy image with the same noise pattern.

Algorithm 3 Sampling

```

1:  $x_T \sim$  testing set (noisy)
2: for  $t = T, \dots, 1$  do
3:   Compute  $\alpha_{t-1} = \left(\frac{t-1}{T}\right)^{\exp}$ 
4:    $x_{t-1} = (1 - \alpha_{t-1})S_\theta(x_t, t) + \alpha_{t-1}x_t$ 
5: end for
6: return  $x_0$ 
```

2.2.3 Example: Diffusion U-Net

In the paper [17], Yang et al. also propose a training procedure for a model called diffusion U-Net using a formulation similar to the one described above. They do not train the network on all possible time steps but instead sample uniformly from the time step range and repeat the procedure until convergence. They also integrate some variations, such as using the Charbonnier loss for a more stable performance [46], and propose an improved sampling algorithm that is less sensitive to errors.

$$L_{\text{char}}(x, y) = \sqrt{(x - y)^2 + \epsilon^2} \quad (15)$$

Using the Diff U-Net model, Yang et al. obtain state-of-the-art results on both the SIDD (Smartphone Image Denoising Dataset) and DND (Darmstadt Noise Dataset) benchmarks [47, 48].

2.2.4 Generalization

On paper the proposed architecture could be applied to all diffusion neural networks by simply adding a time tensor as input and training with the mentioned procedure. This should enable denoising methods to better respond to different noise profiles and potentially improve performance.

Table 3: PSNR and SSIM results on the SSID and DND benchmarks [47, 48]

Methods	SSID Benchmark		DND Benchmark	
	PSNR↑	SSIM↑	PSNR↑	SSIM↑
CBDNet [49]	30.78	0.801	38.06	0.942
VDN [50]	39.28	0.956	39.38	0.952
DANet [51]	39.47	0.957	39.58	0.955
MIRNet [52]	39.72	0.959	39.88	0.956
CycleISP [53]	39.52	0.957	39.56	0.956
MPRNet [54]	39.71	0.958	39.80	0.954
NBNet [55]	39.75	0.959	39.89	0.955
GMSNet [41]	39.63	0.956	40.15	0.961
Uformer [44]	39.77	0.959	39.96	0.956
SwinIR [41]	39.77	0.958	40.01	0.958
Diffusion U-Net	39.81	0.959	40.22	0.962

However, it is possible that this may not be so simple with modern state of the art neural networks such as KB3NET as the self-attention module of the transformer architecture could cause some complications [37]. Further research is needed in this field. In this thesis, I apply the Diffusion U-Net architecture [17] on the CUBS dataset [13] in section 3.2.3 and show its benefits over the traditional U-Net in medical imaging.

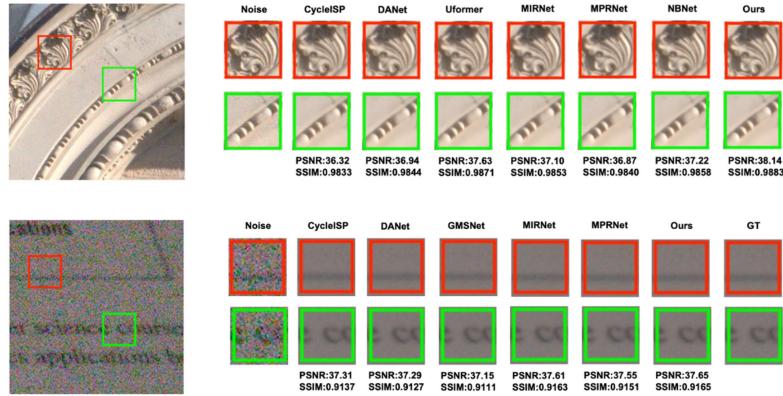


Figure 10: Comparison of denoised images with state of the art supervised methods on the DND dataset (top) and SSID dataset (bottom). Diffusion U-Net is reported as Ours in the images, GT is Ground Truth.

3 Applications in medical imaging

In this section, I propose a pipeline that integrates both unsupervised and supervised denoising diffusion denoising methods to obtain high quality denoised samples from medical images without a clean image to learn from. This is the common case in medical imaging techniques, as it is often very difficult, due to technical limitations, to obtain high-resolution images, especially when it comes to magnified images of critical areas [8] Medical image denoising is a fundamental process in the field of medical imaging, aimed at enhancing the quality of images obtained from various imaging modalities such as MRI, CT scans, and X-rays . These images often suffer from different types of noise due to limitations in imaging hardware, low radiation doses, or fast acquisition times, which can obscure important anatomical details and hinder accurate diagnosis.

In this thesis, I will use ultrasound scan images. Ultrasound imaging is a widely used diagnostic technique that employs high-frequency sound waves to produce images of the inside of the body. This imaging method is particularly valued for its ability to visualize soft tissues. The most prevalent noise pattern in ultrasound imaging is speckle noise, which arises from the interference of reflected waves from small scatters within the tissue [11, 9]. Speckle noise also presents a multiplicative component meaning it is not uniform but will instead vary its profile based on the tissue type.

There are other noise types present in ultrasound scans that might degrade the image. Electronic noise originates from the components of the ultrasound machine, manifesting as random background patterns that obscure details. Thermal noise, also known as Johnson-Nyquist noise, is generated by the random thermal motion of electrons within the device's electronics, contributing a continuous low-level background noise. Cluster noise comes from echoes produced by structures outside the main ultrasound beam which introduce unwanted signals and reduce image contrast and clarity [56].

3.1 Issues with medical imaging

Denoising techniques can often be employed to remove this noise while preserving essential features, ensuring that clinicians can make more reliable interpretations. How-

ever supervised denoising techniques general face the problem of lacking high-resolution clean images to learn from. Additionally, different imaging techniques and organs exhibit very different noise patterns and a model trained on one will likely generalize very badly on to the others [8]. For this reason, an algorithm that can adapt to the noise profile of the image would be ideal.

There is a critical trade-off between noise reduction and preservation of fine details and structural integrity. Some techniques might be very effective at reducing the noise but might lose detail in the process while others might preserve more detail but not get rid of noise in its entirety. Effective denoising techniques aim to minimize noise while retaining important anatomical features and avoiding artifacts that could obscure or distort the true image. Balancing these factors is essential to enhance image quality without losing crucial diagnostic information.

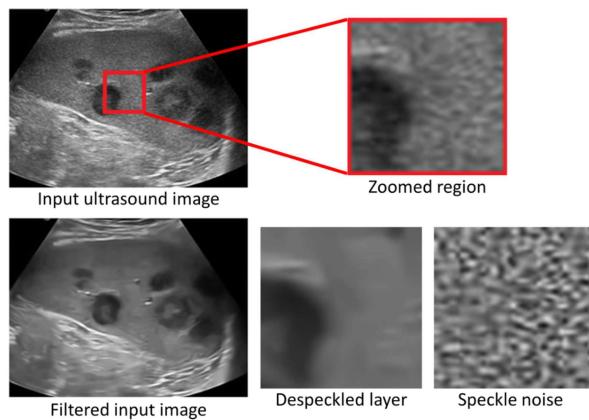


Figure 11: A clinical ultrasound image corrupted with speckle noise (top). The despeckled and speckle noise layers after denoising (bottom) [57].

3.1.1 Noise approximation

When dealing with noise in medical imaging, accurately modeling the noise mathematically is often challenging. Ultrasound images, for example, are affected by various types of noise, including speckle, electronic, thermal, and cluster noise which can combine together. Additionally Speckle noise, being multiplicative, impacts different tissues in varying ways [56, 10]. BM3D is a preferred method in medical imaging due to its adaptability to different noise profiles.

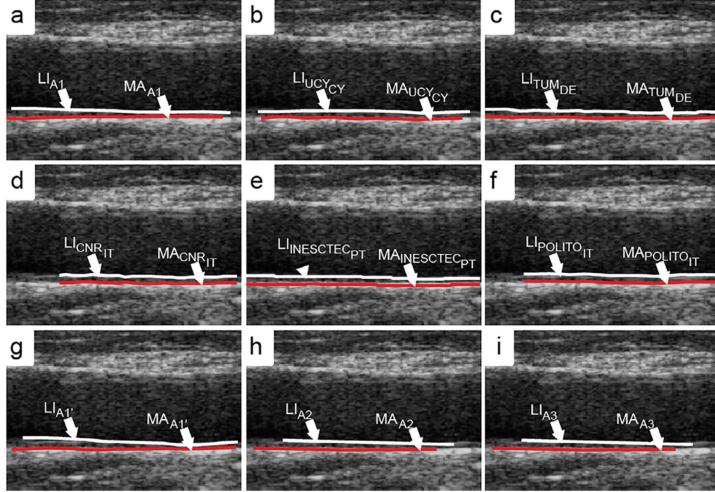


Figure 12: Example of manual and computerized segmentation results on carotid ultrasound scan from CUBS dataset. The lines represent different CIMT (Carotid Intima-Media Thickness) measurement methods [13]

It performs denoising on similar patches, ensuring that the process occurs within tissues of similar characteristics. This approach effectively mitigates issues arising from the multiplicative nature of speckle noise. In the following methods, the cumulative noise profiles in medical images is approximated by using Gaussian noise with a fixed noise level σ [58].

3.2 CUBS dataset: Carotid Ultrasound Boundary Study Dataset

The Carotid Ultrasound Boundary Study (CUBS) dataset is a comprehensive collection of ultrasound images and associated clinical data aimed at analysing and improving computerized intima-media thickness (CIMT) measurement systems [13, 14]. This dataset includes images from 1088 participants, acquired from both sides of the neck, resulting in a total of 2176 images. The data were collected across multiple sites, including villages in Cyprus and a hypertension outpatient clinic in Pisa, Italy. The dataset features manual reference tracings performed by expert analysts and computerized segmentation from five different research groups. It provides a valuable resource for the development and validation of automated CIMT measurement methods and is made publicly available to facilitate further research in this field.

The goal of the dataset is to develop reliable CIMT, carotid intima-media thickness,

measurement methods to estimate the risk of atherosclerosis and prevent it. Accurate denoising is required as the granular nature of the speckle noise degrades image quality and complicates the accurate measurement of Carotid Intima-Media Thickness (CIMT). In this case the most important task is identifying the underlying shapes edges in the image and preserving the continuity of the edges.

3.2.1 Unsupervised denoising methods results on CUBS dataset

To tackle the problem, I tried applying the different unsupervised denoising methods mentioned in Section 2.1.1. Figure 13 shows the results obtained on the CUBS dataset for two sample carotid scans. BM3D with $\sigma = 50$ seems to be the best denoising method in removing the noise profile from the image while also accurately retaining the contours of the shapes. In section 3.2.2, I apply the diffusion BM3D procedure to see if the diffusion procedure can help in preserving the finer details while also increasing the sharpness of the denoised image.

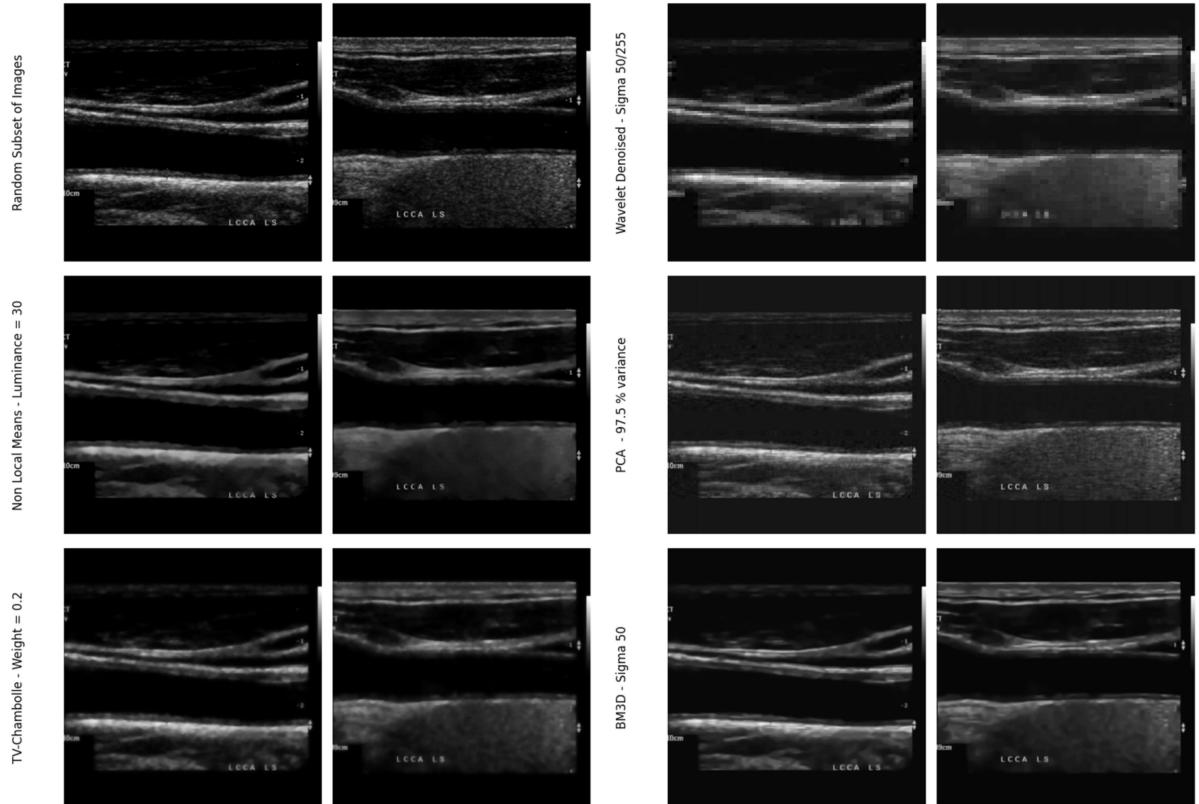


Figure 13: CUBS images denoised by different unsupervised denoising algorithms

3.2.2 Finding the clean distribution: diffusion BM3D

I apply the diffusion BM3D procedure to see if the diffusion procedure can help in preserving the finer details while also increasing the sharpness of the denoised image as it did in the BSD68 dataset images. In Figure 14, I report the resulting images from applying Diffusion BM3D with three noise levels: 1 (the standard BM3D), 10 and 20. Increasing the number of steps, actually seems to increase the quality of the denoised images. Applying more steps results in sharper edges, which can be beneficial in developing CIMT measurement methods.

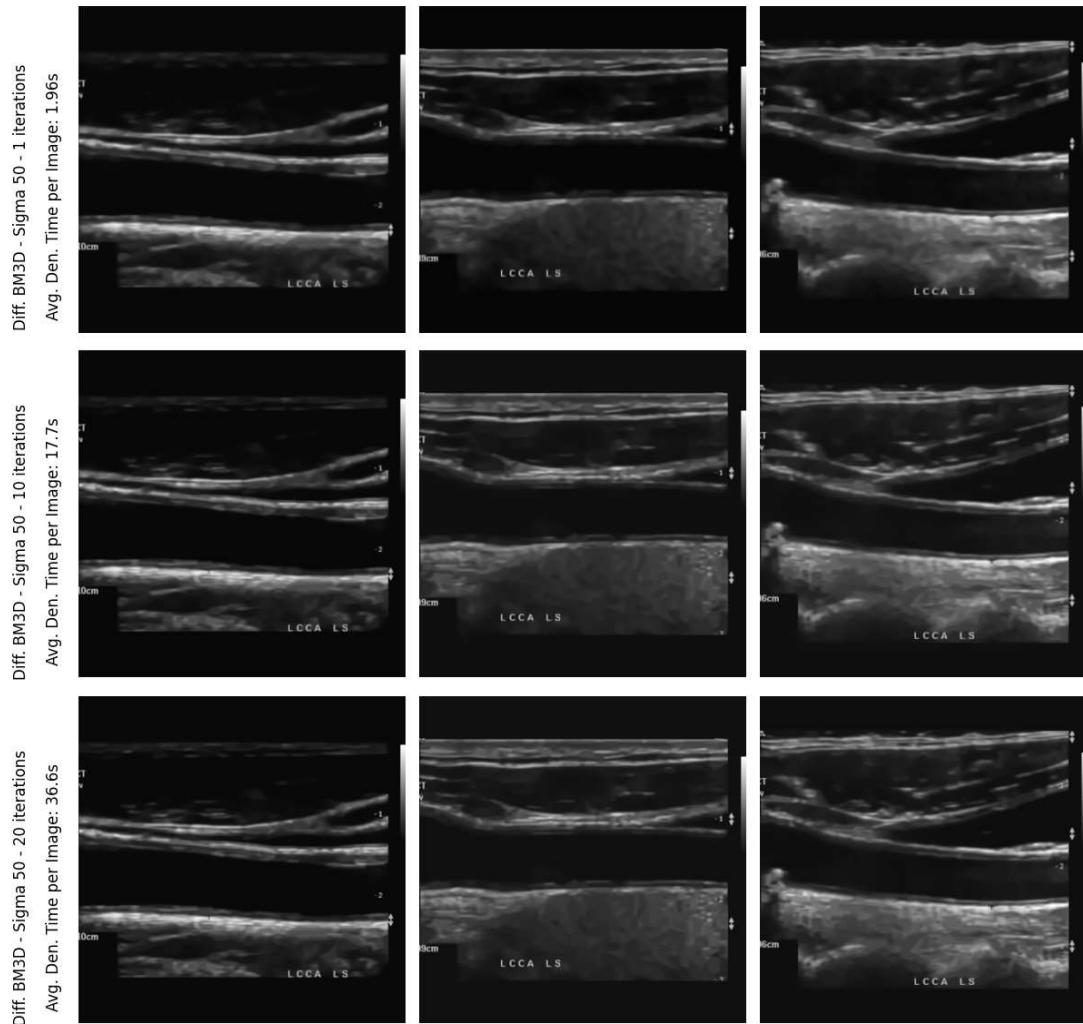


Figure 14: Diffusion BM3D results on CUBS images for different number of steps. The avg. denoising time is reported on the left.

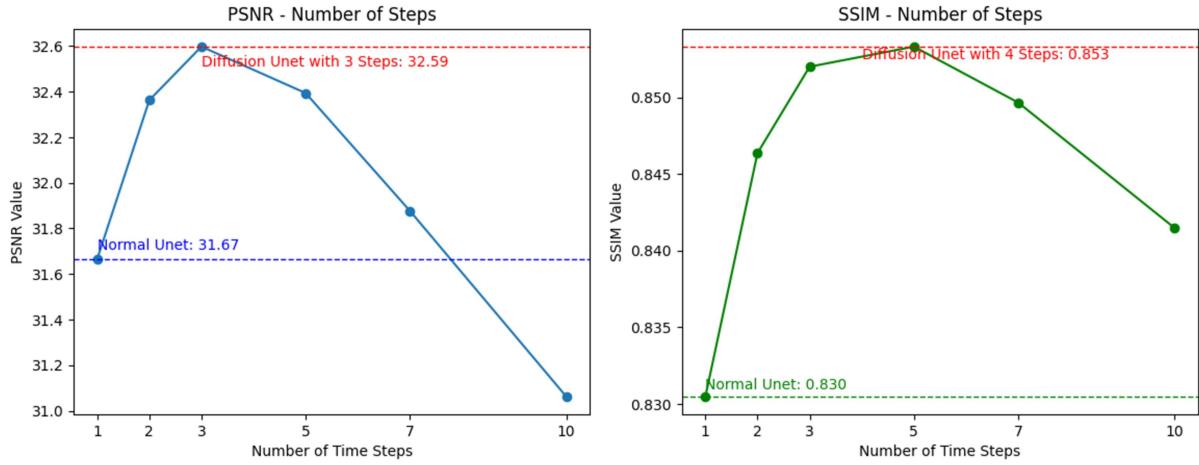


Figure 15: PNSR (left) and SSIM (right) values for different number of steps in the Diffusion U-Net procedure. Values for the baseline U-Net model (blue, green) and optimal number of steps (red) are reported.

3.2.3 Learning the clean distribution: diffusion U-Net

The primary problem with Diffusion BM3D is the significant computational time required to denoise the images. While increasing the number of iterations enhances performance, it also linearly increases the time needed to produce a clean image. This computational process must be repeated for each new image, unlike neural network training that is done once upfront.

An alternative approach involves training a supervised method architecture that can learn from the distribution of denoised images generated by BM3D and generalize to unseen data with much faster sampling times. For this task, I employ the diffusion U-Net architecture proposed by Yang et al. that I discussed in Section 2.2.3. [17].

The raw ultrasound scans from the CUBS dataset are used as noisy images and the denoised predictions of diffusion BM3D with 10 steps are used as clean samples. The data is then split in two, with some images being left out for testing purposes. The goal is to see how well the proposed diffusion U-Net network can learn the denoising procedure applied by Diffusion BM3D and what the ideal number of steps is. The comparison is done in terms of PSNR and SSIM, two metric discussed in Section 2.1.4.

As seen in Figure 15, the PSNR and SSIM values increase significantly by using the diffusion procedure compared to the one-step U-Net.

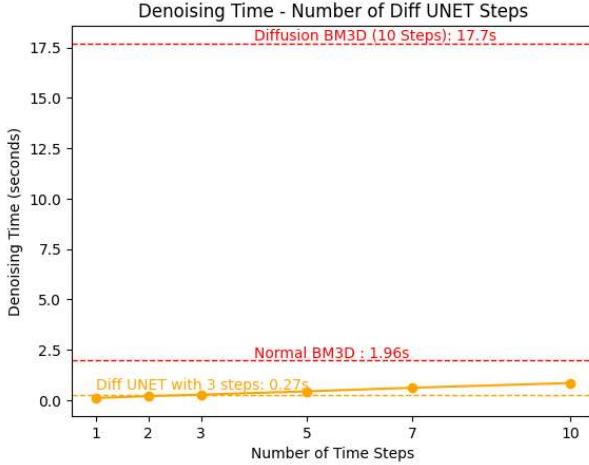


Figure 16: Average denoising time for sampling procedure of Diff U-Net (yellow), BM3D and Diff. BM3D (red)

After three or four iterations however the performance of the diffusion U-Net seems to start to decline. It is possible that this is attributable to the sampling procedure deviating from the target image when using a large number of iterations. A way to enhance this could be implementing a training and sampling procedure to improve the robustness of the iteration or an algorithm to prevent large deviations.

The ideal number of steps for diffusion U-Net in this task seems to be between 3 and 4. The diffusion model with 3 steps increases the denoised image PNSR by 0.92, from 31.67 to 32.59. The diffusion model with 4 steps increases the SSIM by 2.3%, from 0.830 to 0.853. The sampling times for diffusion U-Net are reported in Figure 16. The average sampling time for Diff U-Net with 3 steps is 0,27s compared to 1,96s for the normal BM3D denoising procedure and 17.7s for Diffusion BM3D with 10 steps.

3.2.4 Medical dataset denoising pipeline

I propose the following denoising pipeline for large datasets of medical images exhibiting a consistent noise profile, such as the CUBS dataset. Initially, a sufficiently large subset of available images is denoised using a computationally intensive algorithm like Diffusion BM3D with 10 iterations. Subsequently, a supervised neural network, in this case Diffusion U-Net, is trained on these denoised images to learn and replicate the denoising process of the unsupervised algorithm.

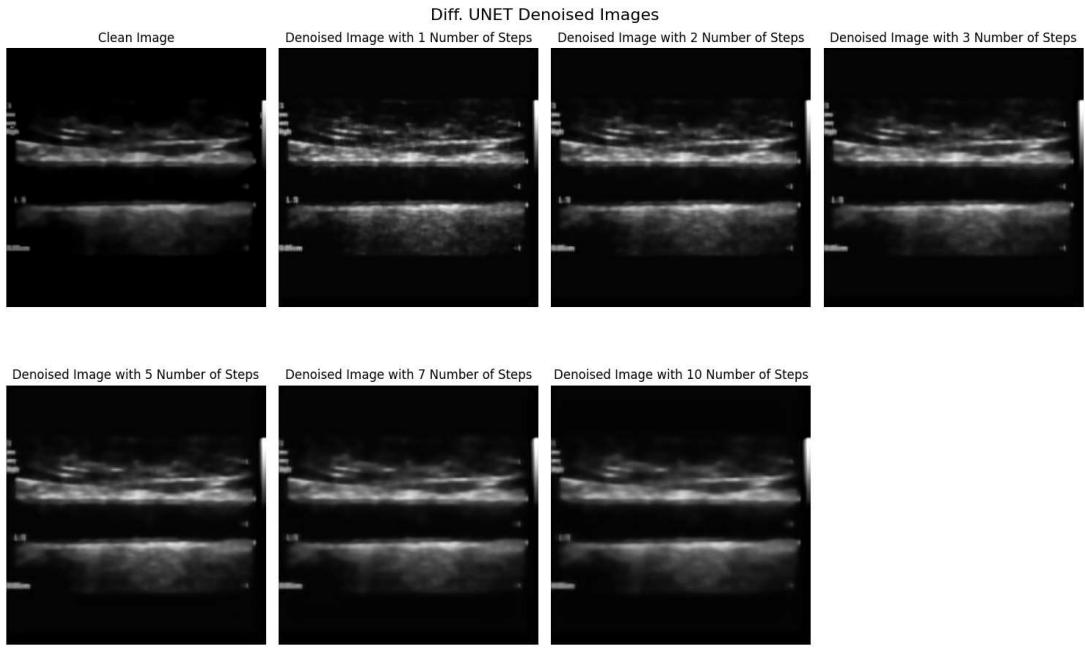


Figure 17: Diffusion U-Net “learned” clean distribution for different number of steps. The clean image is obtained via Diff. BM3D with 10 steps.

Once trained, this supervised neural network can efficiently process new sets of images with similar characteristics, rapidly generating clean samples. This approach requires the initial computational expense to train a model that significantly reduces the time required for subsequent denoising tasks, making it ideal for large-scale medical image datasets.

4 Conclusions

In this thesis work, I illustrated new unsupervised and supervised image denoising methods based on diffusion processes. The main contributions of my thesis can be summarized as follows:

- I proposed a generalized training and sampling procedure for neural networks and algorithms dealing with denoising, which allows for increased performance at the cost of higher training times.
- I designed an improved version of the unsupervised algorithm BM3D and demonstrated that it achieves state-of-the-art results on the BSD68 dataset for unsupervised algorithms. On the BSD68 sigma25 dataset, the Diffusion BM3D achieved a 3% increase in PNSR and a 6% increase in SSIM compared to the standard BM3D algorithm. It achieved similar results on the other noise levels.
- I designed a denoising procedure for computationally efficient medical imaging processing to remove noise from ultrasound scans in the absence of a clean image and demonstrated its potential on the CUBS dataset. This was done by training a supervised neural network, Diffusion U-Net, on the images denoised by an unsupervised algorithm, Diffusion BM3D. The trained Diffusion U-Net achieved a 98% reduction in denoising time compared to Diffusion BM3D denoising, while maintaining a Structural Similarity (SSIM) of 85%.

References

- [1] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, 2020.
- [2] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.
- [3] Robin San Roman, Eliya Nachmani, and Lior Wolf. Noise estimation for generative diffusion models. *arXiv preprint arXiv:2104.02600*, 2021.
- [4] Shoufa Chen, Pei Sun, Yibing Song, and Ping Luo. Diffusiondet: Diffusion model for object detection. *arXiv preprint arXiv:2211.09788*, 2022.
- [5] Arpit Bansal, Eitan Borgnia, Hong-Min Chu, Jie Li, Hamideh Kazemi, Furong Huang, Micah Goldblum, Jonas Geiping, and Tom Goldstein. Cold diffusion: Inverting arbitrary image transforms without noise. *arXiv preprint arXiv:2208.09392*, 2022.
- [6] Hyungjin Chung, Eun Sun Lee, and Jong Chul Ye. Mr image denoising and super-resolution using regularized reverse diffusion. *arXiv preprint arXiv:2203.12621*, 2022.
- [7] Alper Güngör, Salman UH Dar, Şaban Öztürk, Yilmaz Korkmaz, Gokberk Elmas, Muzaffer Özbeş, and Tolga Çukur. Adaptive diffusion priors for accelerated mri reconstruction. *arXiv preprint arXiv:2207.05876*, 2022.
- [8] Sameera V. Mohd Sagheer and Sudhish N. George. *A review on medical image denoising algorithms*, volume 61, page 102036. 2020.
- [9] Rajesh Patil and Surendra Bhosale. Medical image denoising techniques: A review. 4:21–33, 01 2022.
- [10] Jayanta Kumar Dutt and Ranjit Jagtap. Impact of speckle noise on medical image quality and diagnostic performance. *International Journal of Biomedical Imaging*, 2015:1–10, 2015.

- [11] Inderpreet Kaur and Lalit K. Singh. Review of de-speckling algorithms for medical imaging applications. *Biomedical Signal Processing and Control*, 55:101646, 2020.
- [12] Marc Lebrun. An Analysis and Implementation of the BM3D Image Denoising Method. *Image Processing On Line*, 2:175–213, 2012.
- [13] Kristen M. Meiburger, Guillaume Zahnd, Francesco Faita, Christos Loizou, Catarina Carvalho, David Steinman, Lorenzo Gibello, Rosa Maria Bruno, Francesco Marzola, Ricarda Clarenbach, Martina Francesconi, Andrew Nicolaides, Aurelio Campilho, Reza Ghotbi, Efthyvoulos Kyriacou, Nassir Navab, Maura Griffin, Andrie Panayiotou, Rachele Gherardini, Gianfranco Varetto, Elisabetta Bianchini, Constantinos Pattichis, Lorenzo Ghiadoni, José Rouco, and Filippo Molinari. Carotid ultrasound boundary study (cubs): an open multi-center analysis of computerized intima-media thickness measurement systems and their clinical impact. *Ultrasound in Medicine and Biology*, 47:1604–1621, 2021.
- [14] Kristen M. Meiburger, Francesco Marzola, Guillaume Zahnd, Francesco Faita, Christos Loizou, Nolann Lainé, Catarina Carvalho, David Steinman, Lorenzo Gibello, Rosa Maria Bruno, Ricarda Clarenbach, Martina Francesconi, Andrew Nicolaides, Hervé Liebgott, Aurelio Campilho, Reza Ghotbi, Efthyvoulos Kyriacou, Nassir Navab, Maura Griffin, Andrie Panayiotou, Rachele Gherardini, Gianfranco Varetto, Elisabetta Bianchini, Constantinos Pattichis, Lorenzo Ghiadoni, José Rouco, Maciej Orkisz, and Filippo Molinari. Carotid ultrasound boundary study (cubs): Technical considerations on an open multi-center analysis of computerized measurement systems for intima-media thickness measurement on common carotid artery longitudinal b-mode ultrasound scans. *Computers in Biology and Medicine*, 144:105333, 2022.
- [15] Christian Soize. A nonparametric representation of random processes in a diffusion stochastic system. *Probabilistic Engineering Mechanics*, 15(3):277–294, 2000.
- [16] Xavier Amatriain. Transformer models: an introduction and catalog. 02 2023.
- [17] Cheng Yang, Lijing Liang, and Zhixun Su. Real-world denoising via diffusion model, 2023.

- [18] Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2043–2052, June 2021.
- [19] Swati Rai, Jignesh S. Bhatt, and S. K. Patra. An unsupervised deep learning framework for medical image denoising. 2021.
- [20] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A non-local algorithm for image denoising. *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, 2:60–65, 2005.
- [21] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1-4):259–268, 1992.
- [22] David L Donoho and Iain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–455, 1994.
- [23] Ian Jolliffe. *Principal component analysis*. Springer, 2002.
- [24] Michael Elad and Michal Aharon. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing*, 15(12):3736–3745, 2006.
- [25] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on image processing*, 16(8):2080–2095, 2007.
- [26] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019.
- [27] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th international conference on Machine learning*, pages 1096–1103, 2008.

- [28] Norbert Wiener. *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, 1949.
- [29] Jae S Lim. *Two-Dimensional Signal and Image Processing*. Prentice Hall, 1990.
- [30] Dong Wang, Jia Xu, and Ke Xu. An fpga-based hardware accelerator for real-time block-matching and 3d filtering. *IEEE Access*, PP:1–1, 07 2020.
- [31] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proc. 8th Int'l Conf. Computer Vision*, volume 2, pages 416–423, July 2001.
- [32] Quang Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008.
- [33] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.
- [34] Papers With Code. Bsd68 dataset, 2021.
- [35] Mikhail Papkov and Pavel Chizhov. Swinia: Self-supervised blind-spot image denoising with zero convolutions. *arXiv preprint arXiv:2305.05651*, 2023.
- [36] Bo Wu, Yong Ma, Wei Yang, and Zhenwei Shi. Adl: Anomaly detection learning in denoising applications. *IEEE Transactions on Image Processing*, 30:5806–5821, 2021.
- [37] Yi Zhang, Dasong Li, Xiaoyu Shi, Dailan He, Kangning Song, Xiaogang Wang, Hongwei Qin, and Hongsheng Li. Kbnet: Kernel basis network for image restoration, 2023.
- [38] Pengju Liu, Hongzhi Zhang, Wei Lian, and Wangmeng Zuo. Multi-level wavelet convolutional neural networks. 2019.
- [39] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, July 2017.

- [40] Jingwen Su, Boyan Xu, and Hujun Yin. A survey of deep learning approaches to image restoration. *Neurocomputing*, 487:46–65, 2022.
- [41] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1833–1844, 2021.
- [42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [43] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, volume 27, pages 2672–2680, 2014.
- [44] Jiaming Wang, Yulun Zhang, Yapeng Tian, and Yun Fu. Uformer: A general u-shaped transformer for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, 2021.
- [45] Wei Chen, Xiaoming Zhang, Xiangyu Zhang, and Jian Sun. Kb3net: Multi-scale and attention enhanced transformer for image denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2022.
- [46] Pierre Charbonnier, Laure Blanc-Féraud, Gilles Aubert, and Michel Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.
- [47] Zainy M. Malakan, Saeed Anwar, Ghulam Mubashar Hassan, and Ajmal Mian. Sequential storytelling image dataset (ssid), 2023.
- [48] Tobias Plotz and Stefan Roth. Benchmarking denoising algorithms with real photographs. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1586–1595, 2017.

- [49] Shi Guo, Zifei Yan, Kai Zhang, Wangmeng Zuo, and Lei Zhang. Toward convolutional blind denoising of real photographs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1712–1722, 2019.
- [50] Zongsheng Yue, Hongwei Yong, Qian Zhao, Lei Zhang, and Deyu Meng. Variational denoising network: Toward blind noise modeling and removal. In *Advances in Neural Information Processing Systems*, volume 32, 2019.
- [51] Zongsheng Yue, Hongwei Yong, Qian Zhao, Lei Zhang, and Deyu Meng. Dual adversarial network: Toward real-world noise removal and noise generation. In *European Conference on Computer Vision*, pages 41–57, 2020.
- [52] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *European Conference on Computer Vision*, pages 492–511, 2020.
- [53] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2696–2705, 2020.
- [54] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Multi-stage progressive image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14821–14831, 2021.
- [55] Bowen Cheng, Jiangtao Xie, Qian Zhao, Deyu Meng, and Wangmeng Zuo. Nbnnet: Noise basis learning for image denoising with subspace projection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1921–1930, 2021.
- [56] Pierre Gravel, Gilles Beaudoin, and Jacques de Guise. A method for modeling noise in medical images. *IEEE transactions on medical imaging*, 23:1221–32, 11 2004.

- [57] Xin Zhu and Peyman Milanfar. A non-local low-rank framework for video denoising. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4194–4202. IEEE, 2017.
- [58] Ahmed Alnuaimy, Aqeel Jawad, Sarah Abdulkareem, Firas Mustafa, Svitlana Ivanchenko, and Toliupa Serhii. Bm3d denoising algorithms for medical image. pages 135–141, 04 2024.