ECTI Transactions on Computer and Information Technology
Journal homepage: https://ph01.tci-thaijo.org/index.php/ecticit/
Published by the ECTI Association, Thailand, ISSN: 2286-9131

Single Image Denoising through Downsampling and Self-Resolution Restoration Learning

Asavaron Limsuebchuea[1], Rakkrit Duangsoithong[2], Pornchai Phukpattaranont[3] and Terry Windeatt[4]

ABSTRACT

Image denoising using supervised learning effectively removes image noise by learning from available data. However, it may lack efficiency when faced with insufficient data, such as in the case of single images or blind noise. This challenge has led to the adoption of unsupervised learning methods, which utilize the inherent properties of noise to extract and enhance image features. This research aims to leverage the benefits of the downsampling effect for noise removal, even though downsampling may impact image features. Therefore, deep learning must be used to restore image details lost during downsampling. This research proposes the Noisy Low-Resolution to Noisy Super-Resolution (NLR2NSR) framework, which leverages image downsampling to simultaneously reduce image and noise features. A super-resolution network is then used to restore the image features. Experimental results show that under conditions where noise features are less prominent than image features, the NLR2NSR can effectively remove noise and preserve image features using only noisy data for training. However, the NLR2NSR has limitations in handling high-level noise.

Article information:

Keywords: Single-image Denoising, Super-resolution, Self-learning, Image Downsampling, Blind Noise

Article history:
Received: July 4, 2024
Revised: September 19, 2024
Accepted: November 7, 2024
Published: November 23, 2024 (Online)

DOI: 10.37936/ecti-cit.2025191.257424

1. INTRODUCTION

Noise and low-resolution images can significantly impair visual quality and hinder image analysis and processing. Image restoration, an essential process for enhancing image quality [1-3], involves noise elimination and resolution improvement, making images suitable for tasks like object recognition, categorization, and other applications [4] [5]. These issues arise in various phases of image processing due to factors like challenging image acquisition conditions and compression. Noise interference can occur during digital image transmission, and image processing procedures may further degrade image quality. Image interpolation methods increase resolution by adding more pixels, which can lead to edge distortion and noise-related issues. Denoising techniques are employed before interpolation to mitigate these problems. In super-resolution tasks, which involve extending image resolution beyond its original size, noise poses a significant challenge. Hence, selecting appropriate interpolation and denoising methods is crucial to achieve high-quality results in super-resolution [6] [7].

Unsupervised deep learning [8] [9] is widely used in single image and blind noise reduction. However, the training model requires at least two images: one as input and another as the validation target for training the network. Therefore, it is necessary to generate an extra validation set. Another approach involves modifying the learning network structure using techniques like dropout [10] or image-related information to produce specific convolution kernels [11]. However, these methods only capture the local features of a single image.

Image downsampling is the process of reducing the size of an image by skipping or minimising some parts of the data, which operates similarly to a blurring filter. Both downsampling and using a filter can reduce the noise component in the image by filtering out low signal levels. Downsampling can be done by selecting the high-level element in the image, such as using max pooling, which preserves the essential elements.

[1,2,3]The authors are with the Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla, Thailand, Email: 6010130022@psu.ac.th, rakkrit.d@psu.ac.th and pornchai.p@psu.ac.th
[4]The author is with the Department of Electronic Engineering, University of Surrey, United Kingdom, Email: t.windeatt@surrey.ac.uk
[2]Corresponding author: rakkrit.d@psu.ac.th

At the same time, a blurring filter adjusts and reduces some pixel components, resulting in a blurred image compared to downsampling.

This research aims to leverage the benefits of attempting to downsample noisy images, which not only diminishes image features such as edge features but also reduces certain aspects of noise features. While many studies in image super-resolution often focus on enhancing the detail of edge features in images, when confronted with noisy images, the restoration of edge features tends to exacerbate noise features, following the Weiner convolution theory [12-14]. However, considering noise in images as a low-level component compared to edge features, image downsampling consequently reduces noise more significantly than it affects the image's edge features. The behaviour of noise feature reduction through downsampling is, therefore, beneficial for noise elimination in images.

The main objective of this research is to eliminate noise in images by leveraging the benefits of image downsampling processes while preserving the image's edge features. This research proposes the Noisy Low-Resolution to Noisy Super-Resolution (NLR2NSR) framework, which utilizes the ResNet [15] model as the restoration network for edge feature restoration. The NLR2NSR framework can restore a single image without requiring ground truth for validation. Additionally, this research examines the effects of repeated restoration within a learning loop to enhance the quality of the restored image further. The contributions of this research include:

1. Demonstrating the benefits of up- and downsampling in noise removal in images and restoring edge features through the utilization of deep learning techniques,
2. Introducing a self-learning framework with a feedback loop, where the output feeds back into the input, to simultaneously denoise and enhance the edge of the image without using any ground truth for validation training.

Utilizing the benefits of the image downsampling process gives the proposed framework characteristics like joint image denoising and deblurring. This approach is necessary because the proposed framework must remove noise and restore edge features that have been diminished simultaneously. However, this research only employs the downsampling process to reduce noise.

2. LITERATURE REVIEW

In general, for traditional image denoising [16] [17], a kernel is typically chosen to filter out noise signals, which are considered low-level components compared to the edge features of an image. Finding the best parameters to remove noise signals and preserve edge features can be challenging, and many researchers have attempted to use adaptive methods for parameter selection. Traditional image denoising and super-resolution techniques often yield lower performance than deep learning techniques due to the difficulty of selecting appropriate parameters.

In deep learning, supervised learning methods are typically trained using input pairs and ground truth validation images. For example, in many image-denoising tasks [18-20], noisy images are used as input to learning the mapping to the noise-free ground truth. When noise parameters are estimable, deep learning methods often outperform traditional techniques. The network learns to transfer and map the noisy domain inputs into the noise-free domain output by adjusting the weight training parameters through a loss function and optimization process [21]. In cases with a small training dataset, data augmentation [22] [23] is often used to increase the amount of training data, but this can lead to overfitting [24] [25] as the network is still learning the features of the existing images, resulting in redundancy in the dataset. Although some image-denoising approaches attempt to understand the properties of noise rather than trying to map noisy input to a noise-free output, supervised learning often performs poorly in a single image and blind noise scenarios in real-world data.

In real-world scenarios, the primary challenge is to address the problem of single image and blind denoising in the images [26] [27]. The noise level in an image is affected by various factors, such as the surrounding environment and electrical signals in the device, which may change over time. Additionally, when capturing an image using a camera and zooming in on it, the level of noise and resolution can vary, potentially leading to a loss of detail in the edges of the image. Therefore, solving the challenges of single-image denoising and super-resolution is essential for addressing many real-world applications.

Unsupervised learning methods are widely used to address real-world problems caused by insufficient data for learning noise properties. In image denoising, most unsupervised learning algorithms manipulate the network structure or generate new sets of image data based on existing image datasets. For example, many research studies such as Noise2Void [28], Self2Self [10], Noise2Self [11], and Complex-valued deep CNN [29] use convolution or network manipulation techniques to adjust the network weight and remove noise. In Recorrupted-to-Recorrupted [30], a type of data manipulation, a noise generator model is used to create new training pair sets that are then used to train the network. Another approach is Noise2Noise (N2N) [31], which trains the network using pairs of noisy images from the same scenario, but it requires different noisy image pairs for training. The goal of data manipulation techniques is to balance the loss of the network and optimize for all pairs of training sets. Overall, unsupervised image denoising aims to average the results from uncertain data to achieve the best possible outcome for all generated data.

Since this research is related to joint image denoising and resolution enhancement, several studies have addressed this challenge. For instance, [32-34] have utilized supervised learning with their respective network structures. However, these approaches face limitations, particularly concerning the scarcity of datasets in a single image and blind noise scenarios. This research introduces the Noisy Low-resolution to Noisy Super-resolution (NLR2NSR) framework, which combines up- and downsampling techniques with additive noise manipulation to enable the restoration network to self-learn. By leveraging these techniques, the proposed framework can simultaneously restore image denoising and resolution enhancement.

3. THEORETICAL BACKGROUND

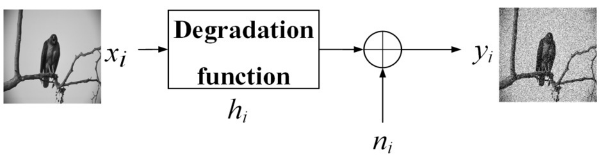

Fig.1: *Block diagram of image distortion caused by joint noise and resolution degradation.*

The image distortion illustrated in Figure 1 refers to the alterations made to an image signal. Typically, the degradation function caused by processing and the environmental noise can distort the image signal.

$$y_i = (h_i * x_i)_{\downarrow} + n_i \tag{1}$$

Equation 1 demonstrates that the image signal x_i is convolved with the degradation function kernel h_i, and the noise n_i is added to the image. Image restoration aims to reconstruct the original high-resolution image by eliminating the impact of noise and damage caused by a damaged image.

3.1 Image restoration using deep learning

In this section, we explain the principles of image edge enhancement. Since this research employs downsampling to eliminate noise and utilizes deep learning to restore image edge features, it parallels training the restoration network to enhance image edges, similar to an image deblurring or a super-resolution task.

Typically, in supervised learning, the goal is to learn how to map inputs from a distorted domain to their ground truth counterparts such that the learning loss approaches minimise value after enough learning cycles. This means that the learning network strives to reduce the difference between the denoised output $(y_{i,pred})$ and the noise-free validation $(y_{j,true})$ in the loss function defined by

$$\boldsymbol{W}_\theta = \underset{\theta}{\operatorname{argmin}} \sum_{i,j=0}^{N} L(y_{i,pred}, y_{j,true}) \tag{2}$$

Where W_θ represents the weights of the learning network with parameters θ that minimize the loss value for all training data pairs $(y_{i,pred}, y_{j,true})$.

If there is not enough training image data available, many unsupervised learning techniques attempt to generate additional validation data from the limited existing dataset $(y_{i,exist})$. This can be achieved by manipulating existing images [30], [31] or modifying the structure of the learning network [10] [11] [28] [29] to create new validation image data $(y_{j,pred})$. The goal is to teach the network how to remove noise from images. This is achieved by assuming that the network aims to average all the noise in the domain to achieve the minimum result, as illustrated in

$$\boldsymbol{W}_\theta = \underset{\theta}{\operatorname{argmin}} \sum_{i=0}^{N} L(y_{i,pred}, \sum_{j=0}^{M} (y_{j,true})) \tag{3}$$

Unsupervised learning is highly beneficial for learning noise reduction since it can decrease the amount of data required for learning. However, quality degradation resulting from image processing remains a significant challenge in image restoration. Unsupervised learning approaches, like the Noise2Noise (N2N) framework [31] and the Recorrupted-to-Recorrupted (R2R) framework [30], operate under the assumption that the noise in the image conforms to a zero-mean distribution. These methods enable the network to indirectly learn this distribution by

$$\mathbf{E}\{y_i | y_j\} \approx \tilde{x}_i \tag{4}$$

Where the estimation $E\{y_i|y_j\}$ refers to taking any given observation of a noisy image y_i and mapping it to another observation of a noisy image y_j within the same observation space. The output of the learning network, denoted as x_i, represents the estimated observation noise-free result. This process can be seen as the network averaging all observations y_i and y_j. In contrast to supervised learning, where the network learns to minimize the difference between the predicted output and the actual target, unsupervised learning optimizes for all possible input observations.

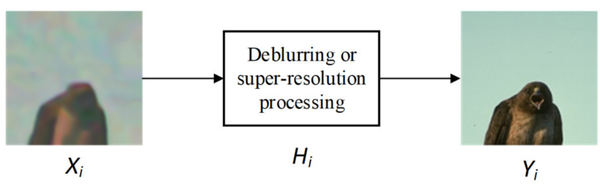

Fig.2: *Utilizing deconvolution for image enhancement to restore clarity and sharpness.*

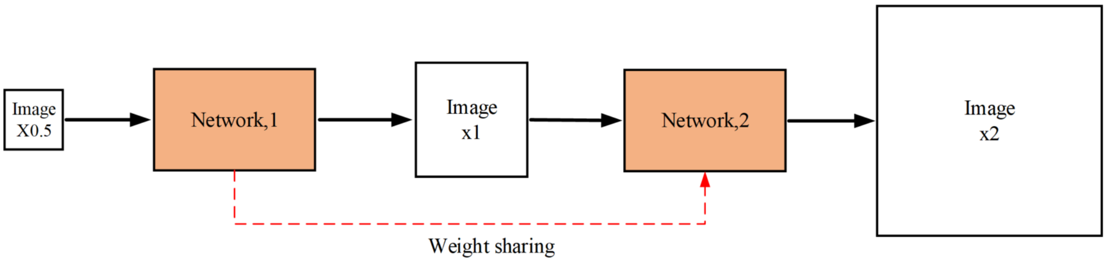

Fig.3: *Zero-short super-resolution (ZSSR) framework.*

In the field of image resolution enhancement, deep learning is employed in Figure 2 to train networks to create deconvolution kernels, aiming to sharpen the edge features of the image. Generally, deep learning-based deblurring and super-resolution techniques rely on training to enhance the dominance of edge features in images. This implies that the adaptive kernel can be viewed as a high-pass filter. In the Fourier domain, the relationship between image distortion can be expressed by

$$\mathbf{Y}_i = \mathbf{H}_i \mathbf{X}_i \tag{5}$$

$$\mathbf{H}_i = \mathbf{D}_S \mathbf{B}_f \tag{6}$$

Where $\mathbf{Y}_i$ refers to the distorted image obtained by convolving the distortion kernel $\mathbf{H}_i$ with the original resolution image $\mathbf{X}_i$. In the case of super-resolution, the distortion kernel $\mathbf{H}_i$ can comprise a blurring kernel $\mathbf{B}_f$ and a downsampling kernel $\mathbf{D}_S$.

Several works [35-37] have demonstrated that image deblurring kernel can be estimated by analyzing the gradient variation between the blurred image and ground truth. In the case of a single image problem, there may be insufficient image data to estimate the blurring kernel and remove noise from the image using Equation 7.

$$\mathbf{X}_i^{'} \approx \frac{\mathbf{H}_i \mathbf{X}_i}{\mathbf{H}_i^{'}} \approx \mathbf{X}_i \tag{7}$$

Where $\mathbf{X}_i^{'}$ represents the deblurred or super-resolution output of the network. $\mathbf{H}_i^{'}$ denotes an estimated kernel provided by the network, and it should be equal to $\mathbf{H}_i$ to optimize the restoration process.

The Zero-Short Super-Resolution (ZSSR) framework [38] shown in Figure 3 is an unsupervised learning approach that leverages accurate blurring kernel estimation to self-learn and successfully increase the resolution of a single image, even though the learning network may be overfit to the image used for self-learning.

$$\{\mathbf{D}_S \mathbf{B}_f\}_{net1} = \{\mathbf{D}_S \mathbf{B}_f\}_{net2} \tag{8}$$

While ZSSR and many other image super-resolution techniques can learn to enhance image resolution on their own, they often overlook the high-level noise component in the image, which can ultimately enhance the noise and negatively impact the system. Noise is a significant factor that must be addressed in the restoration system.

3.2 The effect of noise in learning-based convolution

Deep learning-based image denoising is similar to adaptive kernel methods such as image deblurring and super-resolution. Weiner convolution theory [12-14] explains the impact of noise in convolution and noise suppression using Weiner deconvolution. This method can estimate the deconvolution kernel for removing noise, as shown in

$$\mathbf{X}_i^{'} = \frac{\mathbf{H}_i \mathbf{X}_i}{\mathbf{H}_i^{'}} \left[\frac{1}{1 + \frac{NSR_i}{|\mathbf{H}_i^{'}|^2}} \right] \tag{9}$$

$$NSR_i = \frac{|\mathbf{N}_i|^2}{|\mathbf{X}_i|^2} \tag{10}$$

Where $\mathbf{Y}_i$ is the noisy image $(\mathbf{X}_i + \mathbf{N}_i)$.$\mathbf{X}_i^{'}$ is the denoised image obtained from the deconvolution process $\mathbf{H}_i^{'}$.NSR_i is noise to signal ratio, which can be calculated from the noise-free ground truth $\mathbf{X}_i$ and the noise estimation $\mathbf{N}_i$. However, in scenarios with a single image and blind noise, it is challenging to determine the noise parameters and the original signal from the ground truth. As a result, it can be challenging to learn or create a denoising kernel without a ground truth.

From Equation 10, if $\mathbf{N}_i < \mathbf{X}_i$, where the noise signal is considered as low-level component compared to the image signal, the NSR_i term becomes negligibly small. This allows the learning process to perform effectively, approaching an ideal solution. However, if $\mathbf{N}_i > \mathbf{X}_i$, or if the noise signal behaves as a high-level component relative to the image signal, the learning kernel may not be as effective for image restoration.

The Wiener deconvolution demonstrates that the residual noise in the image resulting from downsampling can be amplified by the convolution process of deblurring or super-resolution processes. This residual noise may be accentuated and could lead to deteriorated results in image edge enhancement.

4. THE PROPOSED NOISY LOW-RESOLUTION TO NOISY SUPER-RESOLUTION (NLR2NSR) FRAMEWORK

Fig.5: *The effects of image down- and upsampling on noisy images, which can reduce noise and distort the edges of the image.*

According to Wiener convolution theory, as the noise level increases, convolution may become less effective at restoring image features. Additionally, it can amplify the high-level components of the noise signal. Therefore, this research will set a condition to consider only cases where the noise remains a low-level component to address the problem effectively.

Image downsampling is a form of convolution that filters out low-level components from an image, which can also reduce the noise component, as seen in the noise reduction effect in Figure 4. The technique of downsampling to reduce noise has been utilized in the SIFT (Scale-Invariant Feature Transform) algorithm [39]. In practice, Gaussian filtering may also be employed to find robust features within an image. However, it distorts the edge features in the image, causing the image to appear blurry. From the observed noise reduction due to the downsampling of an image, this research applies this behaviour to noise elimination together with the restoration of diminished edge features. In this research, the ResNet model [15] is implemented to enhance the resolution of images.

Noise reduction using the downsampling technique resembles employing a convolution filter, which can eliminate low-level noise in the image. However, it also distorts image features. Therefore, this research proposes using the ResNet 16-layer model in Figure 5(A) for self-learning to restore edge features. Since the ResNet model is a CNN-based architecture that preserves feature dimensions within the model, it is well-suited for learning and restoring image details. Although some noise may persist after downsampling, due to the low-level feature nature of noise compared to image features, the noise amplification rate is lower than image features. Additionally, by repeating the learning process and feeding the output back as input, which is then downsampled again, the noise component is significantly reduced. By appropriately adjusting the downsampling scale to match the noise level in the image, noise reduction can be effectively achieved.

The NLR2NSR framework, as depicted in Figure 5(B), resembles a combination of the unsupervised N2N framework for noise removal by adding slight noise to manipulate the training data and the ZSSR framework for restoring edge features with a weight-sharing strategy. The NLR2NSR divides its operation into 2 phases. In phase 1, the restoration network

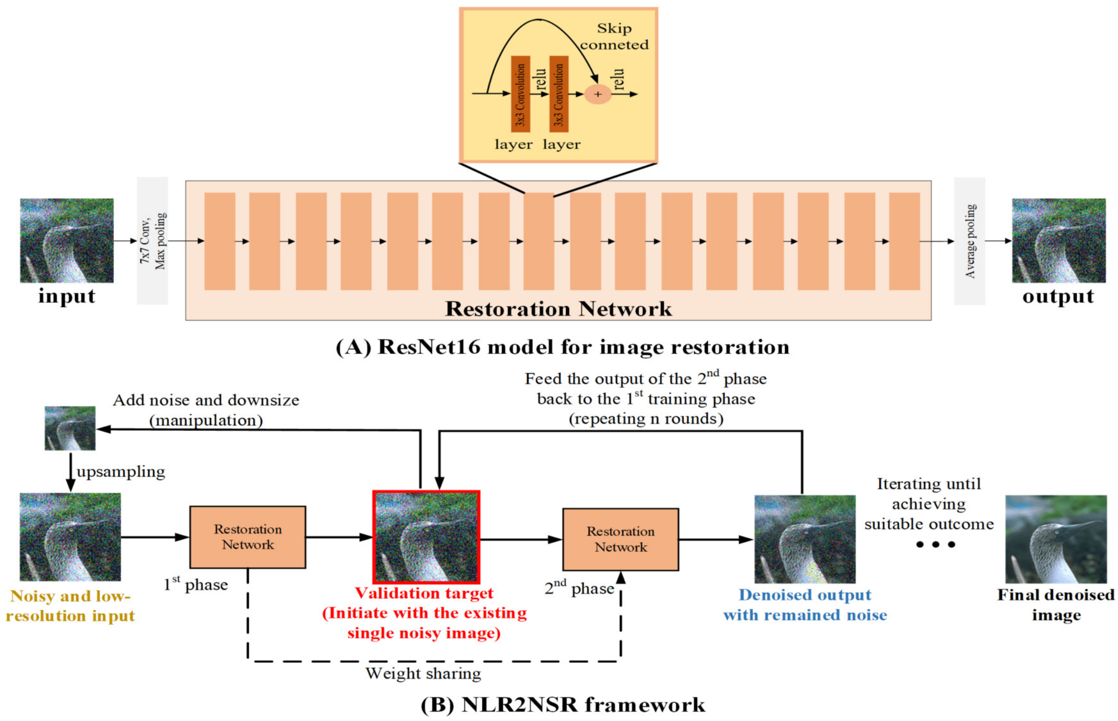

Fig.4: *The proposed NLR2NSR framework for image denoising.*

is trained by specifying the number of iterations for training the restoration network. A single noisy image is downsampled to create a noisy, low-resolution input version, and this single noisy image is used as the validation target after training for the specified number of iterations. In phase 2, the restoration network is employed to remove noise, using the single noisy image as input to eliminate noise. The NLR2NSR framework defines a reiterative process to enhance noise removal results.

The NLR2NSR framework operates in two phases, as illustrated in Figure 5. In the first phase, the framework generates training input by downsampling and upsampling the available noisy image to its original size, thereby self-generating training data. This method reduces the impact of noise, particularly when the noise signal is considered a low-level component, resulting in less visible noise in the image. Furthermore, this approach introduces differing noise characteristics between the available noisy image used as training data and the generated noisy image input, enabling the restoration network to average out the noise signal during the learning process. The network learns to transform the generated noisy image input into the available noisy image used as validation data, continuing this process until convergence is achieved. Once the learning process in the first phase is complete, the second phase uses the available noisy image as input for the trained network, which then outputs a denoised image. The NLR2NSR framework can iteratively use the output from phase two to progressively restore additional image details.

5. EXPERIMENT ON IMAGE DENOISING AND SUPER-RESOLUTION USING NOISY LOW-RESOLUTION TO NOISY SUPER-RESOLUTION (NLR2NSR) FRAMEWORK

This research utilizes the ResNet16 model, which incorporates a 3x3 convolutional kernel with average pooling and ReLU activation functions between the hidden layers. For the input layer, a 7x7 convolutional kernel is employed to extract key image features. During training, the Adam optimizer was used with a learning rate of 0.001. The batch size was set to a single image per batch, as it simulates a scenario with a single image. To evaluate the efficiency of noise reduction and restoration achieved through downsampling, the images are downscaled at three different levels: x2, x4, and x8. This section focuses on restoring real-world noisy images using the NLR2NSR framework.

The image dataset is cropped to 512x512 pixels and then downsampled to 256x256 (for the x2 ex-

Table 1: *The PSNR and SSIM results of the single noisy image of the NIND dataset on x2, x4 and x8 downsampling.*

NIND dataset		**R0**	**R1**	**R2**	**R3**	**R4**	**R5**	**R6**
x2	PSNR	**19.85 ±2.86**	18.22 2.50	15.64 ±2.35	13.12 ±1.97	11.08 ±1.44	9.65 ±0.97	8.71 ±0.65
	SSIM	**0.23 ±0.11**	0.17 ±0.08	0.11 ±0.06	0.07 ±0.04	0.04 ±0.03	0.03 ±0.02	0.02 ±0.02
x4	PSNR	**21.82 ±2.42**	20.96 ±2.63	20.25 ±2.62	19.17 ±2.59	17.93 ±2.49	16.68 ±2.31	15.47 ±2.07
	SSIM	**0.32 ±0.11**	0.28 ±0.10	0.25 ±0.09	0.22 ±0.09	0.18 ±0.08	0.15 ±0.07	0.12 ±0.06
x8	PSNR	22.82 ±4.20	25.48 ±3.11	**26.46 ±3.33**	25.98 ±3.46	25.09 ±3.55	24.18 ±2.58	23.34 ±3.59
	SSIM	0.49 ±0.15	0.55 ±0.11	0.68 ±0.12	0.73 ±0.14	0.75 ±0.16	0.75 ±0.18	**0.75 ±0.20**

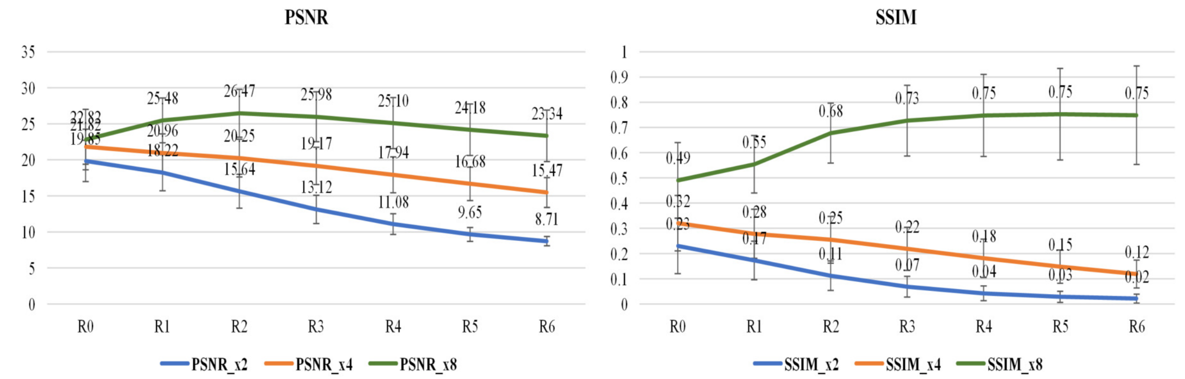

Fig.6: *PSNR and SSIM comparisons of the NIND dataset downsampled by x2, x4, and x8 in multiple rounds.*

Table 2: *The PSNR and SSIM results of the single noisy image of the SIDD dataset on x2, x4 and x8 downsampling.*

SIDD dataset		R0	R1	R2	R3	R4	R5	R6
x2	PSNR	**21.42 ±1.18**	19.54 ±1.06	16.77 ±1.10	14.19 ±1.09	12.20 ±0.90	10.93 ±0.66	10.13 ±0.46
	SSIM	**0.21 ±0.05**	0.15 ±0.03	0.10 ±0.02	0.07 ±0.02	0.05 ±0.01	0.04 ±0.01	0.03 ±0.01
x4	PSNR	**25.44 ±1.33**	24.59 ±1.16	24.08 ±1.16	22.95 ±1.17	21.54 ±1.18	19.97 ±1.21	18.33 ±1.25
	SSIM	**0.59 ±0.05**	0.43 ±0.05	0.36 ±0.05	0.31 ±0.05	0.26 ±0.04	0.21 ±0.04	0.16 ±0.03
x8	PSNR	27.24 ±1.35	27.34 ±1.22	**28.65 ±1.18**	27.82 ±1.25	27.14 ±1.30	26.69 ±1.32	26.36 ±1.34
	SSIM	0.47 ±0.05	0.56 ±0.06	0.74 ±0.05	0.79 ±0.04	0.80 ±0.05	**0.81 ±0.05**	0.81 ±0.05

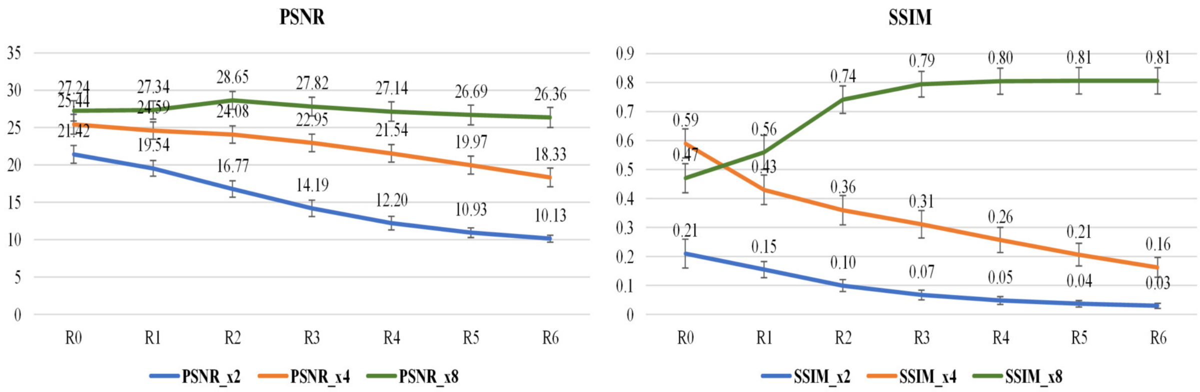

Fig. 7: *PSNR and SSIM comparisons of the SIDD dataset downsampled by x2, x4, and x8 in multiple rounds.*

periment), 128x128 (x4 experiment), and 64x64 (x8 experiment) for the respective downsampling levels. Training is conducted over 100 steps, with 30 iterations per step. In the initial round (R0), the original distorted images are used for training, and shared weights are applied for noise and resolution restoration. The outputs from each round (R1–R6) are fed back into the network every 30 iterations.

The model is evaluated on real-world single-image datasets, specifically NIND [40] and SIDD [41], using 512x512 pixel sub-images for training. Ground truth images are sourced from the lowest ISO settings in the dataset. The results are measured using peak signal-to-noise ratio (PSNR) as the primary metric, alongside the structural similarity index measure (SSIM), to assess the structural quality of the restored images. Table 1 presents the results for the NIND dataset, with the corresponding PSNR and SSIM graphs shown in Figure 6. Similarly, Table 2 shows the results for the SIDD dataset, with PSNR and SSIM graphs in Figure 7.

The results presented in Table 1 and Table 2 indicate that the NLR2NSR framework can generally enhance the PSNR and SSIM outcomes since the first round (R0). Additionally, the restoration outcomes tend to improve further with an increase in the downsampling rate, particularly at x4 and x8. However, when restoring the output image in rounds R1–R6, the x2 and x4 setups consistently produced decreasing PSNR and SSIM values and introduced artifacts in the image. These artifacts were most visible in rounds R4–R6 of the x2 setup of the composite image shown in Figure 8. In contrast, the results of the x8 setup demonstrate that restoring the output image in multiple rounds can lead to an increase in PSNR results. However, this improvement becomes insignificant after a few rounds and drops towards the end. However, the SSIM results steadily improve with each additional round of restoration. These experimental findings suggest that the NLR2NSR framework can effectively restore distorted images affected by noise without needing a ground truth for validation during the learning process.

In the next experiment, we will compare the impact of different noise levels on the NLR2NSR framework. From the previous experiment, we observed both the capabilities and limitations of the downsampling settings in noise reduction. However, insufficient downsampling levels allowed residual noise to affect the learning network. Therefore, in this experiment, clean images from the ground truth of the NIND dataset will be used, and noise will be added

Fig.8: *Sample output of NLR2NSR framework for NIND dataset with x2, x4, and x8 downsampling.*

at levels of $\sigma = 25$, 50, and 95 to evaluate the performance of the NLR2NSR framework in various scenarios.

The experimental results in Figure 9 show that as the noise level increases, the ability of NLR2NSR to remove noise decreases. The resulting images only have increased resolution, which is a consequence of using the super-resolution model for training. This experiment demonstrates that the remaining noise in the images still affects the noise reduction performance of NLR2NSR.

6. COMPARING THE PROPOSED NLR2NSR FRAMEWORK WITH STATE-OF-THE-ART ALGORITHMS

This section will compare the NLR2NSR framework with BM3D [42], which is a filtering-based algorithm, two supervised learning image denoising methods: JDnDmSR [33] and TENet [34], as well as the unsupervised learning approach N2N [31], N2S [11] and SV-N2N [43]. All learning algorithms will be set for training for 50 epochs and 100 steps using a learning rate of 0.001. The BSD300 dataset [44] is utilized for this comparison. The dataset comprises images cropped to 256x256 sizes with Gaussian noise of $\sigma = 50$. The NLR2NSR framework is established by employing x8 downsampling during training and training is set by training for restoring two rounds. Specific comparative results are presented in Table 3, while example images are illustrated in Figure 10. Based on the comparative results, it is evident that the NLR2NSR framework successfully restores noisy images. On average, the results closely resemble those of the N2N algorithm, which is an unsupervised learning method. Although the NLR2NSR achieves slightly lower noise reduction compared to JDnDmSR and TENet, which are supervised learning methods, it can learn noise reduction using a single image through the proposed manipulation process.

7. DISCUSSION

In Section 5, the NLR2NSR framework is introduced, and the results suggest that training with x8 image downsampling at the initial stage (R0) can effectively reduce noise and enhance restoration with more detailed edges compared to x2 and x4 down-

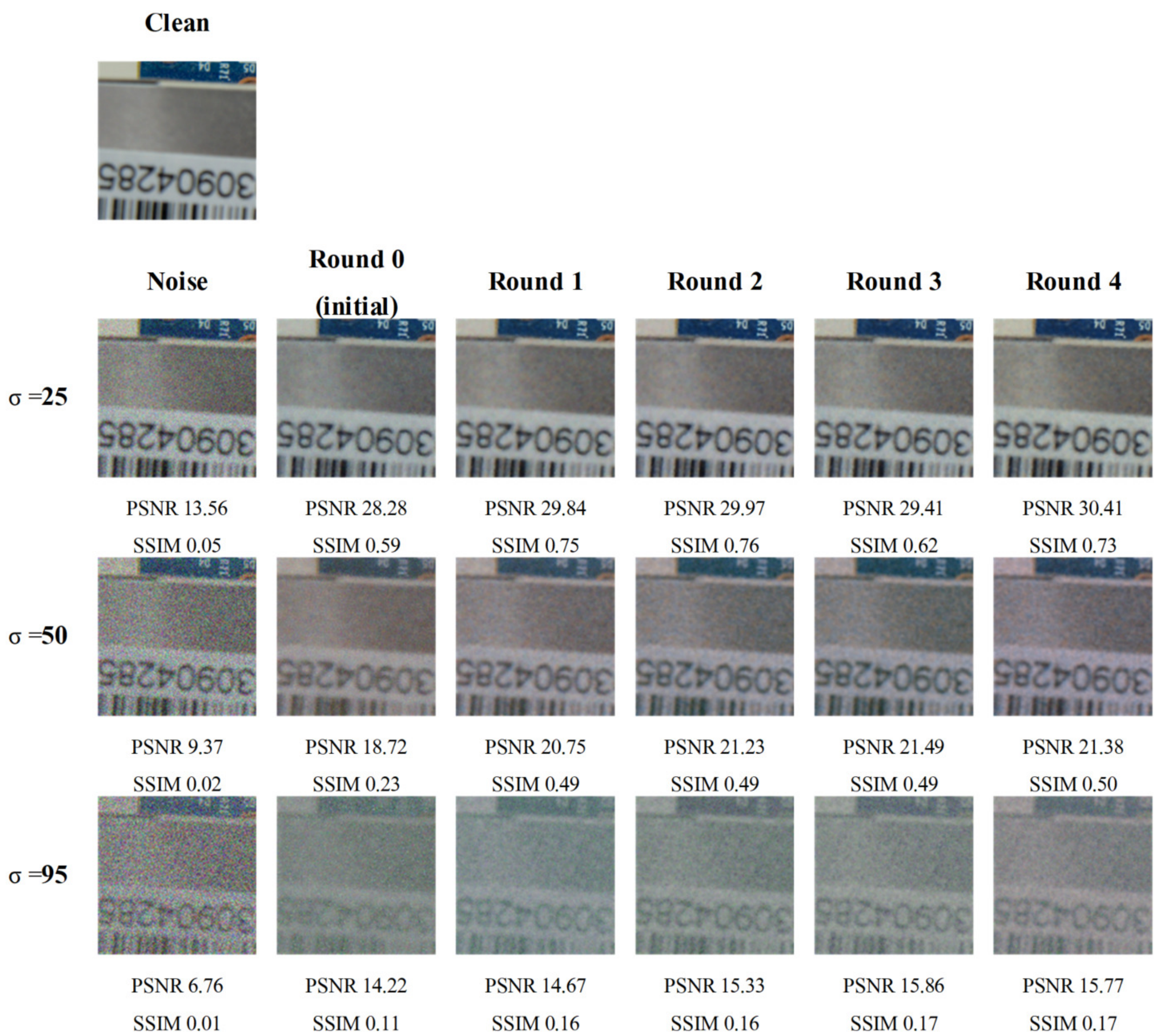

Fig.9: *Comparison of image denoising using the NLR2NSR framework at different levels of Gaussian noise.*

Table 3: *PSNR and SSIM results of the NLR2NSR with restoration algorithms.*

Algorithm	Algorithm type	PSNR	SSIM
BM3D	Conventional filtering	23.60 ±5.23	0.60 ±0.21
JDnDmSR	Supervised learning	26.40 ±4.84	0.72 ±0.14
TENet	Supervised learning	**26.48 ±4.42**	0.72 ±0.12
N2N	Unsupervised learning	25.52 ±2.85	**0.75 ±0.12**
N2S	Unsupervised learning	24.92 ±3.96	0.69 ±0.08
SV-N2N	Unsupervised learning	25.38 ±2.52	0.74 ±0.08
Proposed NLR2NSR	Unsupervised learning	25.56 ±3.67	0.71 ±0.11

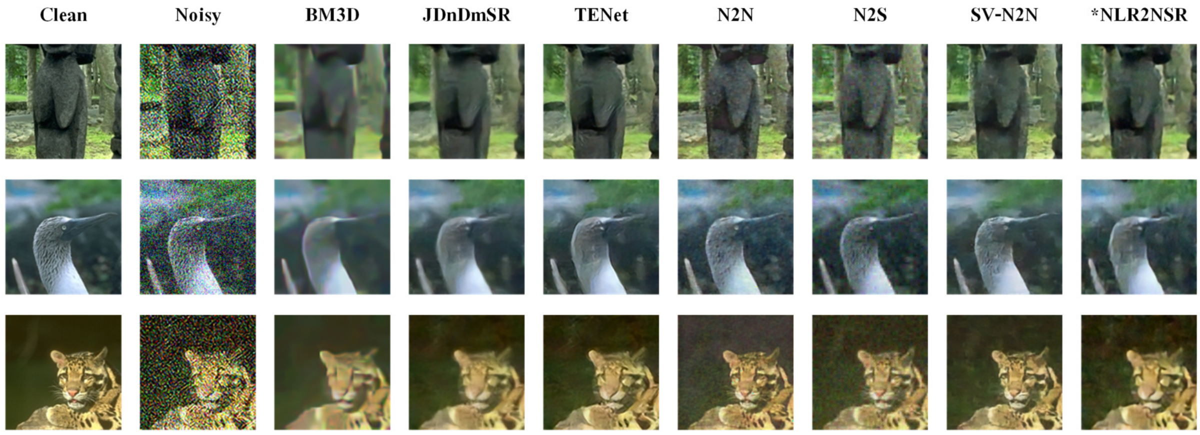

Fig.10: *Example of comparison results of NLR2NSR with other restoration algorithms.*

sampling. The residual noise left behind by lower levels of downsampling (x2 and x4) can be transformed and amplified during the restoration process due to the impact of this residual noise undergoing Wiener convolution in the deep learning convolution process, consequently increasing noise power. Conversely, x8 downsampling is more effective at reducing noise, although it may result in the loss of some edge details. Since the NLR2NSR framework concept is similar to finding robust features like the SIFT algorithm [39], the x8 downsampling level increases PSNR in round 2, as seen in the experimental results in Figures 6 and 7. This sufficiently high downsampling level can reduce noise while preserving essential image features. Preserving these features enables the convolution process during training to amplify them, following Wiener convolution theory.

The proposed NLR2NSR framework uses additive noises injected into a noisy image to generate input and validation pairs for learning restoration. Additionally, the low magnitude of the additive noises compared to the existing noise in a single image dataset, and the presence of remaining noises in the x2 and x4 downsampling setups can negatively impact the performance of resolution enhancement. In contrast, the x8 downsampling setup removes most of the original noise signal, resulting in more effective restoration. The feeding back of the restored output to improve restoration performance can increase PSNR and SSIM results in the x8 downsampling setup. However, excessive feeding back of the restored image can lead to lower PSNR values due to the high-pass filter property of deep learning, causing distortion and artifacts in the image. This effect is particularly evident in the x2 and x4 downsampling setups, as shown in Figures 8 and 9, despite enhancing edge features in the x8 setup with higher SSIM values.

The results shown in Figure 9 demonstrate that different noise levels affect the optimal downsampling settings of the NLR2NSR framework. If the downsampling is too low compared to the amount of noise, such as in the experiment using Gaussian noise with $\sigma = 95$, residual noise will remain in the image. It may enhance unwanted features, leading to incomplete noise removal. However, if the downsampling level is too high, the denoised image may experience distortion of image features, which can also impact the evaluation metrics such as PSNR and SSIM.

The comparison of results in Section 6 reveals that the proposed NLR2NSR framework excels in noise removal and resolution enhancement. Compared to supervised learning, the NLR2NSR framework offers advantages in scenarios where clean images for validation sets are unavailable due to data limitations. Moreover, the training process of the NLR2NSR framework does not require network structure reconfiguration or image augmentation to simulate validation sets, reducing complexity compared to unsupervised learning methods used in the experiment.

8. CONCLUSION

To overcome the challenge of limited training data for restoring noisy images, this research proposes an unsupervised learning approach called the Noisy Low-Resolution to Noisy Super-Resolution (NLR2NSR) framework. The framework combined the concepts of Zero-Short Super-Resolution (ZSSR) and Noise2Noise (N2N) by employing downsampling, upsampling, and noise addition to generate a training set for self-learning. In the first experiment, the restored output was fed back into the restoration process for further enhancement. The results demonstrated that NLR2NSR effectively restored denoising and resolution enhancement by reducing image noise through the downsampling and upsampling processes and enhancing image details through the learning ResNet network. As a result, the proposed framework successfully overcame the limitation of insufficient training data. In the subsequent experiment, by comparing the addition of noise to images at different levels, $\sigma = 25$, 50, and 95, it was shown that the increased noise levels negatively impacted the performance of the NLR2NSR framework. When using a low downsampling level, the effectiveness of the framework decreases as the noise level increases, making it challenging to distinguish image details. This is similar to noise having more prominent or higher-level components than the image features. Consequently, the super-resolution network used in the experiment enhances the resolution of the noise features more than the image features and fails to remove the noise even after multiple iterations. Furthermore, compared with state-of-the-art algorithms, the results demonstrated that NLRSNSR could restore noisy images. However, it showed lower performance in repetitive repairs compared to state-of-the-art algorithms. In future studies, it will be crucial to analyze the relationship between noise intensity and the number of learning iterations, as this factor could significantly impact the restoration of images affected by high-level noise.

ACKNOWLEDGEMENT

This work was supported, in part, by the scholarship awards for master's and Ph.D. studies from Thailand's education hub for ASEAN countries (TEH-AC), grant number TEH2560-015, and, in part, by the Prince of Songkla University for graduate school dissertation funding for the thesis.

AUTHOR CONTRIBUTIONS

Conceptualization, Asavaron Limsuebchuea and Rakkrit Duangsoithong; methodology, Asavaron Limsuebchuea; software, Asavaron Limsuebchuea; validation, Asavaron Limsuebchuea and Rakkrit Duangsoithong; writing—original draft preparation, Asavaron Limsuebchuea; writing—review and editing, Rakkrit Duangsoithong, Pornchai Phukpattaranont, and Terry Windeatt; supervision, Rakkrit Duangsoithong and Pornchai Phukpattaranont. All authors have read and agreed to the published version of the manuscript.

References

[1] L. Fan, F. Zhang, H. Fan, and C. Zhang, "Brief review of image denoising techniques," *Visual Computing for Industry, Biomedicine, and Art*, vol. 2, no. 1, p. 7, 2019.

[2] B. Goyal, A. Dogra, S. Agrawal, B. Sohi, and A. Sharma, "Image denoising review: From classical to state-of-the-art approaches," *Information Fusion*, vol. 55, pp. 220–244, 2020.

[3] H. Chen, X. He, L. Qing, Y. Wu, C. Ren, R. E. Sheriff, and C. Zhu, "Real-world single image super-resolution: A brief review," *Information Fusion*, vol. 79, pp. 124–145, 2022.

[4] Y. Zhou, S. Song, and N.-M. Cheung, "On classification of distorted images with deep convolutional neural networks," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1213–1217, 2017.

[5] M. Koziarski and B. Cyganek, "Image recognition with deep neural networks in presence of noise dealing with and taking advantage of distortions," *Integrated Computer-Aided Engineering*, vol. 24, pp. 337–349, 2017.

[6] W.-C. Siu and K.-W. Hung, "Review of image interpolation and superresolution," in *Proceedings of the 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, 2012, pp. 1–10.

[7] X. Wang, J. Yi, J. Guo, Y. Song, J. Lyu, J. Xu, W. Yan, J. Zhao, Q. Cai, and H. Min, "A review of image super-resolution approaches based on deep learning and applications in remote sensing," *Remote Sensing*, vol. 14, p. 5423, 2022.

[8] D. Zheng, S. H. Tan, X. Zhang, Z. Shi, K. Ma, and C. Bao, "An unsupervised deep learning approach for real-world image denoising," in *International Conference on Learning Representations*, 2021.

[9] X. Wang, Y. Yang, Z. Guo, Z. Zhou, Y. Liu, Q. Pang, and S. Du, "Real-World Image Super Resolution via Unsupervised Bi-directional Cycle Domain Transfer Learning based Generative Adversarial Network," *ArXiv*, vol. abs/2211.10563, 2022.

[10] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2self with dropout: Learning self-supervised denoising from single image," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1887–1895, 2020.

[11] J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," in *Proceedings of the 36th International Conference on Machine Learning*, vol. 97. PMLR, , pp. 524–533, 2019.

[12] J.-C. Yoo and C. W. Ahn, "Image restoration by blind-wiener filter," *IET Image Processing*, vol. 8, no. 12, pp. 815–823, 2014.

[13] H. Chen, Z. Cen, C. Wang, S. Lan, and X. Li, "Image restoration via improved wiener filter applied to optical sparse aperture systems," *Optik*, vol. 147, pp. 350–359, 2017.

[14] S. Trambadia and P. Dholakia, "Design and analysis of an image restoration using wiener filter with a quality-based hybrid algorithms," in *2015 2nd International Conference on Electronics and Communication Systems (ICECS)*, pp. 1318–1323, 2015.

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[16] B. Goyal, A. Dogra, S. Agrawal, B. Sohi, and A.

Sharma, "Image denoising review: From classical to state-of-the-art approaches," *Information Fusion*, vol. 55, pp. 220–244, 2020.

[17] N. You, L. Han, D. Zhu, and W. Song, "Research on image denoising in edge detection based on wavelet transform," *Applied Sciences*, vol. 13, no. 3, p. 1837, 2023.

[18] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[19] K. Bajaj, D. K. Singh, and M. A. Ansari, "Autoencoders based deep learner for image denoising," *Procedia Computer Science*, vol. 171, pp. 1535–1541, 2020.

[20] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Transactions on Image Processing*, vol. 27, no. 9, pp. 4608–4622, 2018.

[21] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[22] C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning," *Journal of Big Data*, vol. 6, no. 60, 2019.

[23] M. Xu, S. Yoon, A. Fuentes, and D. S. Park, "A comprehensive survey of image augmentation techniques for deep learning," *Pattern Recognition*, vol. 137, p. 109347, 2023.

[24] B. Wu, Z. Liu, Z. Yuan, G. Sun, and C. Wu, "Reducing overfitting in deep convolutional neural networks using redundancy regularizer," in *Artificial Neural Networks and Machine Learning – ICANN 2017*, pp. 49–55, 2017.

[25] M. M. Bejani and M. Ghatee, "A systematic review on overfitting control in shallow and deep neural networks," *Artificial Intelligence Review*, vol. 54, pp. 6391–6438, 2021.

[26] P. Yang, J. Sturtz, and L. Qingge, "Progress in blind image quality assessment: A brief review," *Mathematics*, vol. 11, no. 12, 2023.

[27] M. A. N. I. Fahim, N. Saqib, S. K. Siam, and H. Y. Jung, "Denoising single images by feature ensemble revisited," *Sensors*, vol. 22, no. 18, 2022.

[28] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2void - learning denoising from single noisy images," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2124–2132, 2019.

[29] Y. Quan, Y. Chen, Y. Shao, H. Teng, Y. Xu, and H. Ji, "Image denoising using complex-valued deep cnn," *Pattern Recognition*, vol. 111, p. 107639, 2021.

[30] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2043–2052, 2021.

[31] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," *CoRR*, vol. abs/1803.04189, 2018.

[32] Y. Kim, H. Ryu, S. Lee, and Y. J. Lee, "Joint demosaicing and denoising based on interchannel nonlocal mean weighted moving least squares method," *Sensors*, vol. 20, no. 17, 2020.

[33] W. Xing and K. Egiazarian, "End-to-end learning for joint image demosaicing, denoising and super-resolution," in *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3506– 3515, 2021.

[34] G. Qian, Y. Wang, J. Gu, C. Dong, W. Heidrich, B. Ghanem, and J. S. Ren, "Rethinking learning-based demosaicing, denoising, and super-resolution pipeline," in *2022 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–12, 2022.

[35] F. Wang and Y. Li, "Robust kernel estimation for single image blind deconvolution," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 481–484, 2012.

[36] P. Sodhani, A. Bordia, and K. Karthik, "Blind content independent noise estimation for multimedia applications," *Procedia Computer Science*, vol. 54, pp. 549–557, 2015.

[37] Y. Li, C. Liu, X. You, and J. Liu, "A single-image noise estimation algorithm based on pixel-level low-rank low-texture patch and principal component analysis," *Sensors*, vol. 22, no. 22, 2022.

[38] A. Shocher, N. Cohen and M. Irani, "Zero-Shot Super-Resolution Using Deep Internal Learning," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 3118-3126, 2018.

[39] W. Burger and M. J. Burge, "Scale-Invariant Feature Transform (SIFT)," London: Springer London, pp. 609–664, 2016.

[40] B. Brummer and C. De Vleeschouwer, "Natural image noise dataset," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1777–1784, 2019.

[41] A. Abdelhamed, S. Lin and M. S. Brown, "A High-Quality Denoising Dataset for Smartphone Cameras," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, pp. 1692-1700, 2018.

[42] A. A. Yahya, J. Tan, B. Su, M. Hu, Y. Wang, K. Liu, and A. N. Hadi, "BM3D image denoising algorithm based on an adaptive filtering," *Mul-*

timedia Tools and Applications, vol. 79, no. 27, pp. 20391–20427, 2020.

[43] L. Asavaron, D. Rakkrit, and J. Jermphiphut, "A self-validation noise2noise training framework for image denoising," *The Imaging Science Journal*, vol. 72, no. 7, pp. 1–16, 2023.

[44] 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," *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Vancouver, BC, Canada, vol.2, pp. 416-423, 2001.

Asavaron Limsuebchuea received the B.Eng. and M.Eng. degrees in electrical engineering from the Prince of Songkla University, Songkhla, Thailand, in 2014 and 2017, respectively, and received Ph.D. degree in electrical engineering from the Prince of Songkla University, Songkhla, Thailand, in 2024. He is currently a research assistant with the Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University. His research interests include computer vision, deep learning, machine learning, and embedded systems.

Rakkrit Duangsoithong received the B.Eng. degree in electrical engineering from Chiang Mai University, Thailand, in 1995, the M.Eng. degree in electrical engineering from Prince of Songkla University, Songkhla, Thailand, in 2001, and the Ph.D. degree from the University of Surrey, Guildford, U.K., in 2013. He is currently an associate professor with the Electrical and Biomedical Engineering Department, the Faculty of Engineering, Prince of Songkla University. His current research interests include machine learning, computer vision, signal processing, and generative AI.

Pornchai Phukpattaranont received the B.Eng. and M.Eng. degrees in electrical engineering from the Prince of Songkla University, Songkhla, Thailand, in 1993 and 1997, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Minnesota, Minneapolis, MN, USA, in 2004. He is currently a professor of electrical and biomedical engineering with the Prince of Songkla University. Examples of his ongoing research include the pattern recognition system based on electromyographic signals, electrocardiographic signals, and microscopic images of breast cancer cells. His current research interests include signal and image analysis for medical applications and ultrasound signal processing. He is a member of the ECTI Association and Thai Biomedical Engineering Research Societies.

Terry Windeatt received the BSc degree in Applied Science from University of Sussex, followed by M.Sc. in Electronic Engineering from University of California, B.A. in theology and PhD degree from University of Surrey, U.K.. After lecturing in Control Engineering at Kingston University, UK, he went to live and work in the USA for eight years. He worked on Intelligent Systems in the Research and Development Departments of General Motors and Xerox Corporation in Rochester, NY. His industrial R&D experience is in modelling/simulation for intelligent automotive and office-copying applications. He returned from the United States in 1984 to join the Department of Electrical and Electronic Engineering at the University of Surrey, and has lectured in Machine Intelligence and Control Engineering for over twenty years. He has worked on various research projects in the Centre for Vision, Speech and Signal Processing, and his current research interests include Neural Networks, Pattern Recognition, and Computer Vision.