IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 8 May 2024, accepted 12 May 2024, date of publication 20 May 2024, date of current version 28 May 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3402748

RESEARCH ARTICLE

Self-Augmented Noisy Image for Noise2Noise Image Denoising

ASAVARON LIMSUEBCHUEA[1], **RAKKRIT DUANGSOITHONG**[1,2], **AND PORNCHAI PHUKPATTARANONT**[1], **(Member, IEEE)**

[1]Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University, Songkhla 90110, Thailand
[2]Smart Industry Research Center, Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University, Hatyai Campus, Songkhla 90110, Thailand

Corresponding author: Rakkrit Duangsoithong (rakkrit.d@psu.ac.th)

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

ABSTRACT Image denoising is a critical task in image processing aimed at removing noise artifacts. Typically, supervised deep learning often necessitates a large number of pairs of noisy and noise-free images for training. Noise2Noise techniques have demonstrated efficiency in noise removal without relying on a noise-free ground truth. This is achieved through a learning process that approaches input to target points, balancing results across all training inputs. While Noise2Noise can be adapted for single image denoising, it still faces challenges in single image and blind noise scenarios. To address this issue, our research introduces the concept of self-augmented noisy images for self-supervised Noise2Noise single image denoising. The proposed method leverages the behavior of the training process, which strives to balance the loss values appropriately for each training set. By utilizing the same noisy image for both input and validation to learn self-identification, it produces another set of noisy images that mimic the input noisy images. From the experimental results, measured using Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) metrics, it is evident that the proposed self-augmented strategy enables Noise2Noise to remove noise in single image scenarios. Additionally, it achieves performance comparable to other unsupervised denoising methods without requiring additional augmentation manipulations.

INDEX TERMS Image denoising, single image denoising, blind noise, self-supervised, self-augmentation.

I. INTRODUCTION

Image denoising is a crucial task in image processing. Noisy images [1], [2], arise from disruptive signals added to the image data. These signals can originate from sources such as cameras with inherent noise or tools that introduce interference signals into the image. The presence of noise diminishes the clarity of image details, impacting both color and object structure, ultimately leading to a loss of sharpness. Therefore, image denoising [3], [4] is a process that focuses on reducing or eliminating these interfering signals to restore the image to its intended quality, allowing the denoised image to be usable.

The associate editor coordinating the review of this manuscript and approving it for publication was Yi Zhang.

Image denoising can be divided into two types: filtering-based and deep learning-based. Filtering based image denoising [5] often involves using image filtering techniques designed to reduce noise by adjusting the pixel values of the image according to specific formulas or methods. This helps enhance clarity and improve the quality of the image for display or further processing. There are numerous techniques available to enhance images affected by visible noise, addressing aspects such as sharpness and color correlation across various elements of image data. Deep learning [6], [7] is a learnable filter that can effectively outperform filtering methods based on traditional approaches. This is because deep learning utilizes neural networks to automatically learn complex features from data. Deep learning models can adapt well and effectively learn to classify both image features and

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License.
For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

noise feature patterns. Additionally, deep learning can capture intricate relationships within the data, allowing it to effectively differentiate between noise signals and image signals. The learning process in deep learning facilitates automatic feature extraction, making it more robust to different noise scenarios without the need for manually designed filters. As a result, deep learning-based approaches often outperform traditional spatial filtering methods, especially in scenarios with diverse and complex noise patterns.

The conventional approach in deep learning typically involves utilizing numerous pairs of noisy and noise-free images for training, known as the Noise2Clean (N2C) strategy [6], [8]. An alternative, Noise2Noise (N2N) [9], demonstrates that deep learning can directly identify the clean image signal by training the network to transform one noisy image domain into another noisy image domain. This eliminates the need for the network to explicitly learn clean image signals and significantly reduces the required amount of training image data. However, it is crucial that the training images are replicable from the same viewpoints to ensure a sufficient representation of image features for effective learning by the denoising network. Nevertheless, both Noise2Clean and Noise2Noise still encounter data insufficiency issues when learning from blind noise scenarios and single image scenarios. Therefore, the main issue with the Noise2Noise strategy still resembles the supervised Noise2Clean strategy because, by the concept of training, images used for training must have different distributions of noise to ensure that the images retain enough image signal features for the denoising network to recognize. This implies that the Noise2Noise strategy still requires training image data that can be acquired or replicated. However, the challenge with blind noise and single image denoising is the absence of images for validation, making it impractical to apply Noise2Noise straightforwardly.

Deep learning typically involves utilizing a loss function to train the network, aiming to optimize the weighting for all training pairs. However, this method often introduces uncertainties in the training loss, particularly in domains affected by noise interference. Leveraging this uncertainty within the training loss, as proposed in studies on uncertainty loss functions [10], [11], provides an opportunity to generate additional noisy images for augmentation. By pairing these augmented noisy images with available noisy images, denoising networks can be effectively trained using the Noise2Noise strategy. This approach capitalizes on the behavior of learning loss to enhance the denoising process.

This research aims to address the challenge of blind image denoising by introducing a self-augmented noisy image network to Noise2Noise, which still faces difficulties in a blind noise scenario in real world application, such as in tasks involving medical images [12] like CT, MRI and X-rays, as well as images from remote sensing and geographic information systems (GIS) [13]. The proposed self-augmented noisy image network is designed to generate a new set of noisy images from the available noisy image set, following the same principles as the Noise2Noise algorithm. The augmented noisy images share a similar noise distribution with the available noisy images, but the network cannot precisely replicate the input as the output. This characteristic allows the generated noisy image output to be utilized as validation data in training the Noise2Noise network, effectively removing the noise signals. The experimental results demonstrate that the output of the self-augmentation network can effectively serve as validation training for the Noise2Noise strategy.

Overall, this research has two main objectives, which are to present the contributions obtained from the proposed method as follows:

- This research improves the Noise2Noise strategy, enabling it to eliminate noise in single image or blind noise scenarios. This is achieved by generating noisy images that resemble the existing noisy images.
- This research utilizes and explores the potential of self-identification and the uncertainty of the learning process in deep learning to be utilized as an augmentation process with noisy image data for the task of image denoising.

The structure of the manuscript comprises the following sections: In Section II, an overview of related work is provided, discussing the background of image denoising algorithms and the various noise challenges encountered in image processing. Section III describes the theoretical aspects of noise occurrence in images, image denoising algorithms, and the functioning of image denoising algorithms, including conventional filtering-based, Noise2Clean, and Noise2Noise methodologies. Section IV introduces the proposed self-augmented noisy image for Noise2Noise image denoising, explaining the concept and workflow of the algorithm. In Section V, the setup for each algorithm used in the experiments will be described, along with the performance metrics and the dataset employed for testing. Sections VI and VII present the experimental results and discuss the comparison of the noise reduction performance between the proposed method and other image denoising algorithms. Finally, Section VIII summarizes the conclusions and contributions.

II. RELATE WORK

A. IMAGE DEGRADATION PROBLEM IN IMAGE PROCESSING

Image degradation [1], [2], [3], [4] can be classified into two types depending on the impact on the image: convolution degradation and additive signal degradation. The convolution degradation model involves a degradation kernel that is convolved with the image, causing distortion or reducing the sharpness of the image. This type of degradation arises from various causes, including digital image processing algorithms such as image blurring or downsampling, and environmental conditions such as rain or fog. On the other hand, degradation by additive signal often results from interference signals causing abnormal pixel behavior in certain parts of the image, commonly referred to as additive noise. This type of noise [1] can be categorized into various types depending on the environmental conditions. For example, Gaussian noise arises from environmental conditions during image capture or circuit noise within the camera device. Impulse noise typically occurs during data transmission and can result in certain points in the image being affected, displaying either the minimum or maximum value within the range of data values. Poisson noise is closely related to Gaussian noise and often originates from low-light conditions or the detection of photons. Sparkle noise arises from capturing images with light reflecting off uneven surfaces, causing changes in characteristics or wavelength before impacting the camera. This results in irregular noise appearing in the image.

Generally, analysis and determination of the relationships and parameters of these degradation models enable the design of appropriate filter kernels or masks to effectively remove interference signals using conventional filtering methods. However, in many situations, the identification of the degradation model that occurs may not be clear, as noise signals can arise from multiple and simultaneous interference models, making it difficult to analyze the components of the noise. This situation is known as the blind noise problem. Additionally, for single-image problems, insufficient image data may pose another challenge to analysis.

B. IMAGE RESTORATION ALGORITHM

Image restoration algorithms are designed to restore images to have sharpness and eliminate components that degrade the quality of the image. As mentioned, image degradation can be categorized into two types, leading to image restoration algorithms being divided into two problem types as well. These include algorithms used to remove convolution kernel components and algorithms used to remove additive signal components. The algorithm for deconvolution kernel components aims to enhance and restore the details of the image to have sharp edges and object corners, such as image deblurring and image super-resolution [14], [15]. Furthermore, the loss of image feature sharpness can also arise from the environment, such as in works [16], [17], [18], [19], which address the problem of sharpness loss in images in a hazy environment. Additionally, issues in real-world images also lack the data necessary for effective learning. Therefore, in some research, solutions to real-world single image problems are also proposed, as seen in [15], [18], and [19].

This research focuses on image denoising. Initially, conventional kernel image denoising involves designing a filter kernel according to the type and intensity of noise present in the image. This can be categorized into two domains: spatial domain filtering [5], [20], which directly considers the spatial relationships to design a filter kernel that aligns with the type of noise, and transform domain filtering [21], [22], [23], which attempts to separate image and noise components to minimize the impact of noise filtering on the image signal. However, conventional filtering often requires parameter settings like thresholds or the number of components to separate, such as in Principal Component Analysis (PCA) [21] or Wavelet transform [22], [23], making it less user-friendly and prone to errors in parameter tuning. While techniques like estimating noise levels with methods such as the Wiener filter [24] or optimizing points for Total Variation (TV) regularization [25], relying on a single kernel filter for estimation and denoising may not always yield sufficient effectiveness.

Currently, deep learning plays a significant role in image processing, particularly in image restoration tasks. This is because of its complex architecture, which excels in extracting diverse features and allows the filter to adapt itself. Consequently, deep learning filters demonstrate higher accuracy and efficiency compared to conventional filters, without the need for parameter tuning to reduce estimation errors. By learning from datasets specific to the problem domain, various deep learning architectures have been developed. Examples include ResNet [26], which incorporates residual units to prevent overfitting in deep layers, and U-Net [27], which prevents feature loss in an autoencoder-like structure. Both ResNet and U-net are widely applied in image restoration tasks due to their effectiveness learned from datasets. Additionally, the Vision Transformer (ViT) [28] is a structure that emphasizes exploring relationships within sub-windows in its own structure to understand the interconnections of objects in the image. ResNet, U-net, and ViT have all seen widespread development and application in image restoration tasks, employing various techniques to enhance the efficiency of self-attention in the learning process. However, ResNet and U-net often utilize limited datasets, whereas ViT requires a larger dataset to enable the structure to recognize appropriate features effectively.

Noise2Noise (N2N) [9] stands out as an unsupervised learning technique that revolutionizes image denoising, eliminating the need for noise-free ground truth data during training. N2N learn to remove noise from 2 datasets comprising noisy images: one for input and the other for validation. Unlike traditional Noise2Clean (N2C) approaches [6], [8] that aim to converge to a single correct answer through loss function and optimization behavior, N2N introduces a distribution of random noise between input and output. This induces a learning behavior that averages results between the input and output, impacting the trained weights. Consequently, applying these weights to denoise images yields satisfactory results. However, the N2N technique necessitates two sets of noisy images for training, posing challenges in situations that obtaining both a noise-free version and another noisy version may be impractical.

In the realm of unsupervised and self-supervised learning for image denoising, there are two subcategories depending on the nature of image data manipulation. The Recorrupted-to-Recorrupted (R2R) [29] and self-validation Noise2Noise (SV-N2N) methods [30] attempt to generate a new set of noisy images from available noisy images. R2R, as an algorithm of this type, generates two image sets by adding and subtracting random noise into the available image set. This process is utilized for training the denoising network. On the other hand, SV-N2N employs a method of adding small-sized noise to manipulate available noisy images, creating a new set of noisy images for training. The goal is to eliminate noise signals in a manner similar to N2N. The second method involves manipulating the network structure through training. This is done by using a set of available noisy images for both input and validation targets. During training, features that occur within the network structure are perturbed instead. For instance, Noise2Void (N2V) [31] and Noise2Self (N2S) [32] utilize blind-spot kernels to obscure the perception of pixels in a certain portion of the image. This prompts the training network to estimate the missing pixel values, resulting in a reduction of noise levels. Similarly, Self2Self (S2S) [33] closes the perceptual awareness of network pixels during training, employing the dropout technique as a blind spot that can be adjusted. However, both blind-spot and dropout techniques may lead to the loss of image features during perception.

This research introduces a self-augmented noisy image network for self-supervised training the blind noise and single image N2N image denoising framework. The self-augmented network aims to create a noisy image dataset that closely resembles available noisy images for use as self-validation training in the N2N image denoising strategy. This research aims to address the challenges of blind noise and single image denoising using the N2N strategy as a foundation, without the need for many clean or noisy images for training the denoising network. In addition, this research also considers the analysis of the self-augmented network, which is trained using only mean square error loss. It further explores the introduction of using the variance difference between input and output images to enhance the appropriateness of the loss function for self-augmentation learning.

III. THEORY

A. IMAGE DENOISING WITH SPATIAL FILTERING APPROACH

Noise in images is often considered an additional component according to the degradation model [1] by (1).

$$x_i' = h_i * x_i + n_i. \tag{1}$$

where x_i' is the noisy image that arises from the clean image signal x_i, which is disturbed by the noise n_i, and by h_i, which is the degradation kernel.

Filtering methods typically rely on the characteristics of a filter that depend on the type of noise present. There are various techniques that can be employed for noise removal, such as addressing the spatial domain's additional noise component through wavelet transform [22], [23] or PCA [21]. Frequency domain analysis, like FFT-based approaches [34], is also utilized. In addition, there are learning-based algorithms like Total Variation (TV) denoising [25], which adjusts the parameter values of filter kernels to suit the specific type of noise that needs to be removed. However, conventional filters primarily focus on finding optimal parameters that match the noise signal, requiring prior knowledge of the type and level of noise for selecting the most suitable parameters.

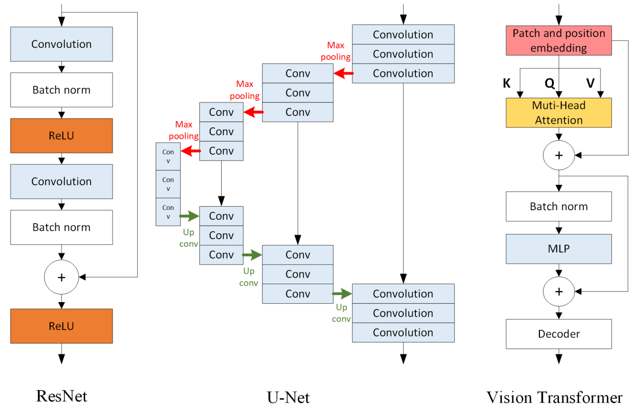


FIGURE 1. Comparison of the basic structures of ResNet, U-Net, and ViT for image denoising.

Block-mathcing 3D (BM3D) filtering [20] is a denoising algorithm based on the concept of grouping sub-images with similar structures before applying noise removal, where the Wiener filter [24] is employed. This approach leverages the inherent nature of images, which often contain sub-components with structures resembling flat regions, edges, and corners, allowing effective grouping and analysis of noise characteristics. By analyzing these sub-images, filter parameters suitable for the specific noise type can be designed. For instance, in the case of Gaussian noise with a mean value of zero, grouping these sub-images and averaging them can be employed to derive an effective noise removal.

Nevertheless, conventional filters involve adjusting parameter values to design filter kernels for noise removal. The obtained results often rely on recognizing the type and level of noise interference. The noise in images may arise from various types and levels of interference, presenting a challenge in blind noise scenarios where filtering kernels may struggle to efficiently address the issue.

B. IMAGE DENOISING WITH DEEP LEARNING APPROACH

This research utilizes three deep learning models for self-augmentation to assess their performance: ResNet [26], U-Net [27], and ViT [28] shown in Fig 1. These models are chosen because each structure has characteristics that uniquely impact image restoration tasks. ResNet is a CNN-like structure that preserves feature counts in each layer but adds skip-connected layers to bring features from previous layers to prevent data loss, making it suitable for tasks such as feature restoration in images. U-Net is an autoencoder-like structure that attempts to extract features from images using max pooling and convolution for upsampling to maintain the output size. It also incorporates skip-connected layers to reuse features from previous layers, similar to ResNet. Although U-Net is initially designed for image segmentation tasks, its

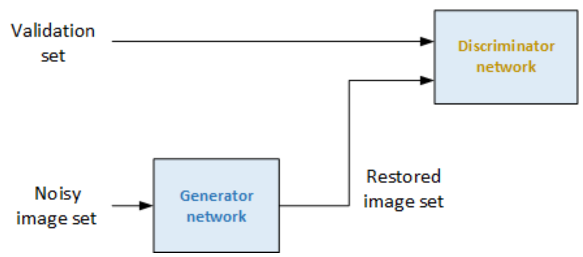


FIGURE 2. The competitive learning structure of GANs for image denoising.

feature reduction structure also makes it suitable for image denoising tasks. Vision Transformer is a structure designed to manage relationships between image patches independently, adding a mechanic beyond learning with input and validation datasets alone. ViT networks can link relationships between image patches using Key (K) and Query (Q) matching with Value (V) during learning, which is suitable for tasks requiring sequential relationships in images, such as image segmentation and classification. For image restoration, ViT can learn the relationships of noise occurring throughout the image patches.

Generative adversarial networks (GANs) in Fig 2 are a type of deep learning architecture that can be used in tasks such as image restoration [14], [35] and image augmentation [36], [37]. The structure of GANs consists of two competitive networks: the generator network and the discriminator network. In image restoration tasks, the generator network attempts to generate a restored image and sends it to the discriminator network for validation against a clean image validation set. During training, the generator strives to create restored images to make the discriminator distinguish between the generated images and the validation set. However, the drawback of GANs is that they easily suffer from overfitting during training due to the competitive nature of their architecture, potentially leading to the generation of artifacts [14].

C. IMAGE DENOISING WITH CONVENTIONAL NOISE2CLEAN LEARNING APPROACH

Deep learning serves as a powerful technique for deep feature extraction, enabling the learning of filter parameter adjustments within its structure through a weight-adjusting mechanism during supervised learning using dataset. The process of learning noise removal through deep learning is referred to as supervised Noise2Clean (N2C). The learning mechanism of deep learning attempts to understand the transition from a noisy input domain to a noise-free target domain, which is defined by a validation set. In the field of image denoising, this can be likened to the endeavor of aligning the noise distribution of noisy pixels to match the target clean pixels, as illustrated in Fig 3.

While deep learning is a highly effective technique compared to filtering-based approaches, it still heavily relies on a substantial amount of data for learning. The issue of blind noise becomes problematic when the image dataset used for learning is insufficient, impacting the efficiency of noise removal.

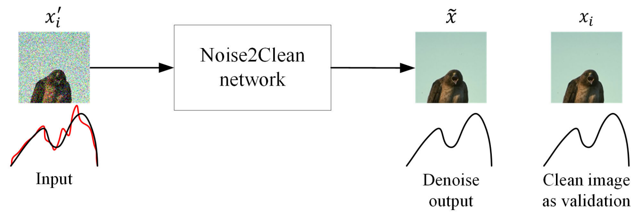


FIGURE 3. Conventional training in Noise2Clean (N2C) involves learning to remove noise by utilizing clean validation training.

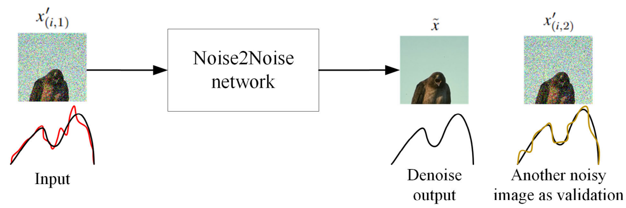


FIGURE 4. Using the Noise2Noise (N2N) training framework involves utilizing noisy images as input and another noisy domain for validation training.

D. IMAGE DENOISING WITH NOISE2NOISE LEARNING APPROACH

Noise2Noise (N2N) is an image denoising learning approach that utilizes pairs of independently generated noisy images to remove noise from images. This method leverages the converging efforts of the loss function during training [9], [38]. In this case, training with independently generated random noise data prompts the network to adjust the gradient of the loss function [39] to fit each set of these random noise data. However, because the validation of noisy image training involves another set of random noise, the network cannot converge to the validation target for every validation set. As a result, the network exhibits behavior analogous to averaging all gradients, causing the randomly distributed noise to appear as an average by equation (2). Considering noise as a zero-mean Gaussian distribution, allowing the network's output to reveal clean image features hidden within the noisy mask.

$$\tilde{x} \sim \mathbb{E}\{x'_i\} = \frac{\sum_{i=1}^{N} x'_i}{N}. \tag{2}$$

where $\tilde{x}$ is the denoising result of the expected observation $\mathbb{E}\{x'_i\}$, which can be estimated from the average value of the noisy image set x'_i, in the case where the noise is random Gaussian distribution.

The N2N averages results through the L_1 or L_2 loss function, making it challenging for deep learning optimization to precisely converge toward the noise distribution of the entire validation set. Consequently, it achieves a balance by adjusting the loss values appropriately for all validation noisy images.

Based on the N2N technique, various methods have been developed to address issues with noisy images that present blind noise problems. These approaches involve manipulating image datasets for both single image datasets, allowing the network to learn without the need for image pairs. This can be categorized into two groups: self-supervised by data manipulation and by convolutional network manipulation.

For the self-supervised by data manipulation approach, the R2R framework [29] introduces a method involving the addition and subtraction of random noise in training images to create image pairs for learning. This allows the network to average the noise values by

$$x'_{(i,1)} = x_i + n_{(i,1)}. \quad (3)$$
$$x'_{(i,2)} = x_i + n_{(i,2)}. \quad (4)$$

On the other hand, the SV-N2N framework [30] adds additive noise to an available noisy image to create a noisy image pair for an existing single noisy image by

$$x'_{(i,2)} = x'_{(i,1)} + n_{(i,2)}. \quad (5)$$

In terms of network structural manipulation [31], [32], [33], it involves attempting to modify the learning network through various methods. In the S2S [33], this is achieved by learning from a single image using dropout in hidden layer of network structure while training with the same noisy image as input and validation. The N2V [31] employs blind spot kernels, and the N2S [32] utilizes random sampling of pixels in the image for calculating loss.

However, data manipulation-based methods still need to estimate values for image manipulation. This is essential for creating a validation set. The use of excessively high noise levels may result in suboptimal noise removal outcomes. Additionally, network structure manipulation-based approaches, including the use of dropout and random blind spots, as well as randomly selected samples, may lead to data loss in each training round, introducing uncertainty in the training loss.

IV. THE PROPOSED: SELF-AUGMENTED NETWORK FOR NOISE2NOISE IMAGE DENOISING

The concept of noisy image augmentation in this research uses the same idea and principle as the N2N strategy [9]. In practice, the loss function in the deep learning system, where the diversity of features in the data is high, makes it impossible for the loss value to completely converge to all training set. Therefore, the learning process not only aims to minimize this loss value but also involves appropriately balancing it for all the data used in learning. Due to this nature of learning in deep learning, even if the same dataset is fed as both input and validation for self-identification learning, the deep learning process cannot perfectly mimic and generate output images identical to the input. However, this behavior of deep learning can be leveraged to simulate new datasets for noisy image augmentation to eliminate noise using the N2N technique.

Therefore, this research experiment involved using a single noisy image set, where the deep learning network attempted

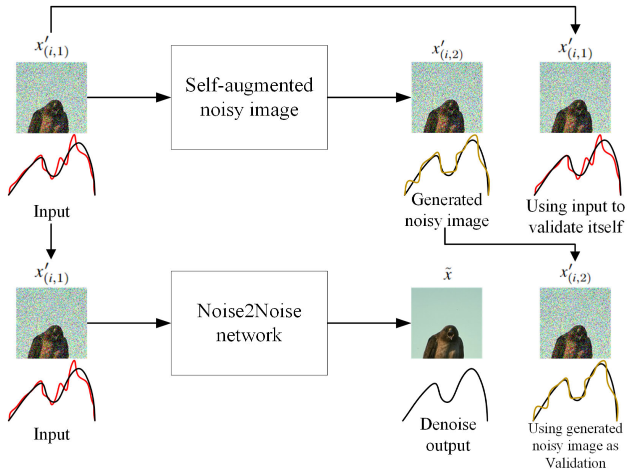


FIGURE 5. Self-augmented network and using the generated noisy image output as noisy image validation for Noise2Noise denoising network.

to self-simulate following the principles of learning through the loss function. However, as per the fundamentals of learning, it is challenging to make every data point in the learning process precisely converge to the validation training set. Therefore, the learning process of self-simulation resulted in an output that closely resembled the input, as guided by the loss function.

The overall methodology of this research, as shown in Fig 5, involves two sub-networks:

- **Self-augmented noisy image network**: This network is tasked with self-augmentation and the generation of noisy images. It takes the original noisy image $x'_{(i,1)}$ set as input and is trained using the same original noisy image $x'_{(i,1)}$ set for validation, producing a set of noisy images $x'_{(i,2)}$ as the output.
- **Noise2Noise denoising network**: In the second step, the generated images ($x'_{(i,2)}$) from the self-augmented network are paired with the original noisy images ($x'_{(i,1)}$) to create a training pair dataset ($x'_{(i,1)}, x'_{(i,2)}$). This dataset is then used to train the Noise2Noise Removing Network, which focuses on effectively removing noise and enhancing overall image quality.

The self-supervised algorithms [29], [30], [31], [32], [33] demonstrate that networks can capture the distribution pattern of noise from a given training dataset. In this research, experiments were conducted to have the network self-simulate, recognizing that the learning process of the network may not be perfect due to the manifold of features in the dataset. This imperfection arises from the network attempting to mimic only the image features and noise distribution it encounters during learning. While conventional deep learning can already be learned to transform noise from one domain to another through mean squared error (MSE) loss functions, this study reveals that relying solely on MSE may lead to overfitting in the self-augmented network.

To address this, the research proposes incorporating the difference in variance between the input and output images

TABLE 1. The comparison of learning characteristics between the proposed algorithm and other image denoising algorithms.

Algorithm	Number of datasets used for training	Learning type	Training description
BM3D [20]	1 image (each image is processed individually)	Conventional filter	No learning, but uses appropriately set parameters
DnCNN [6]	Many images for the best performance	Supervised learning	Uses noise-free validation and noisy input pairs for training
N2N [9]	At least 2 noisy images with the same screen	Unsupervised learning	Creates noisy image pairs for a set from noise-free image dataset
N2V [31]	At least 1 noisy image	Unsupervised learning	Uses an image as both input and validation through blind-spot kernel convolution
S2S [33]	At least 1 noisy image	Unsupervised learning	Uses an image as both input and validation through a network with dropout technique
N2S [32]	At least 1 noisy image	Unsupervised learning	Validates the learning process by randomly sampling from input and output data
SV-N2N [30]	At least 1 noisy image	Unsupervised learning	Manipulates noisy images to create pairs for training
Proposed Self-augmented noisy image for N2N framework	At least 1 noisy image	Unsupervised learning	Uses a learning network to **self-identify** input noisy images and uses the output as noisy image pairs for N2N training framework

as an additional parameter in the loss function during the training of the self-augmented network. This is achieved by utilizing the variance in a normalized parameter format.

$$L_{\text{norm}}(x'_{(i,1)}, x'_{(i,2)}) = \frac{L_{\text{MSE}}(x'_{(i,1)}, x'_{(i,2)})}{\text{var}(x'_{(i,1)}, x'_{(i,2)})}. \tag{6}$$

where $L_{\text{norm}}(x'_{(i,1)}, x'_{(i,2)})$ represents the normalized loss and $L_{\text{MSE}}(x'_{(i,1)}, x'_{(i,2)})$ represents the MSE. $\text{var}(x'_{(i,1)}, x'_{(i,2)})$ is the normalization factor due to the large size of the MSE and variance values; it has an impact on the parameter optimization calculations in the early stages of training. Normally, using L_{MSE} alone allows the network to learn to generate noisy images. However, learning solely through MSE still affects the use of the generated images in this step for denoising with the N2N technique. This issue will be demonstrated and discussed in the experiment section.

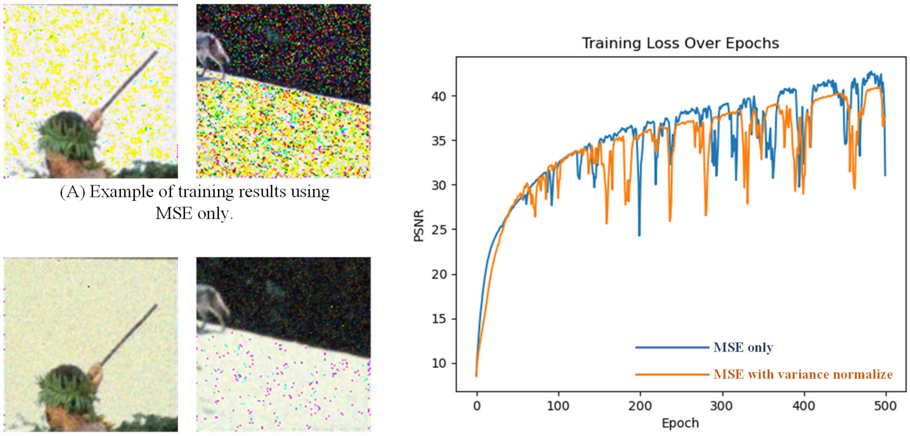
Table 1 compares the characteristics and learning approaches of the algorithms reviewed earlier in image denoising. Our proposed method emphasizes the utilization of single noisy images and operates effectively in blind noise scenarios. By focusing on single image denoising and blind noise reduction, our method enables N2N to learn from generated noisy images instead. This strategy eliminates the need for a large dataset or complex learning models, focusing on the effectiveness of single noisy images in blind noise scenarios. By optimizing parameters and leveraging generated noisy images, our approach streamlines the denoising process while maintaining high performance and computational efficiency in practical applications.

V. EXPERIMENT SETUP AND IMPLEMENTATION DETAIL

In this experiment, the capabilities of various algorithms were tested using the BSD300 [40] and Set14 [41] datasets for standard validation. Gaussian noise with standard deviations of 25 and 50 was added to the images, and the SIDD [42] and NIND [43] real-world noisy images were used as the blind noisy image dataset. All datasets were used in their original aspect ratios but were cropped to a size of 128×128 pixels. The peak signal-to-noise ratio (PSNR) and Structural Similarity Index (SSIM) were used to compare the performance of each image denoising algorithm.

For this experiment, the aim is to compare the proposed method of self-augmented Noise2Noise, utilizing ResNet, U-Net, ViT, and GANs as learning structures. The ResNet used in this experiment is a 16-layer model, while the U-Net model consists of an 8-layer encoder and an 8-layer decoder. ViT is configured to divide the image into 8×8 patches for patch embedding. GANs were employed for noisy image augmentation, with the U-Net structure serving as the generator and a discriminator utilizing a CNN with 8 convolutional layers to evaluate the generated results from the generator. The 9th layer of VGG was utilized for feature extraction. During training, a noisy image set is fed as validation input through the VGG feature extractor before being passed to the discriminator, and it serves as the noisy image input for the generator to enable it to produce a new set of noisy images. In all convolutional layers of the usage network, the convolution was configured with a 3×3 kernel size and strides of 1, utilizing the ReLU activation function. During training, all models utilized a default learning rate of 0.001 with the Adam optimizer from the Keras platform and the MSE loss function. GANs utilized MSE and binary cross-entropy as the content loss and adversarial loss, respectively. Training was set for 100 epochs, with each epoch consisting of 100 steps.

In this experiment, an additional technique has been introduced by utilizing the generated noisy image output from the self-augmented network. Instead of using the available noisy input as the input to the Noise2Noise network and the generated noisy image as the validation set, as explained in the theory section, this research has swapped the input and validation sets. This creates $[x'_{(i,1)}, x'_{(i,2)}]$ as the input and $[x'_{(i,2)}, x'_{(i,1)}]$ as the validation set. This serves as an additional



(A) Example of training results using MSE only.

(B) Example of training results using MSE with variance normaliztion.

(C) Comparing training performance over epochs between using MSE and MSE with variance normalization.

FIGURE 6. The experimental results comparing the usage of only MSE and MSE with normalized variance loss function over training epochs.

augmentation strategy during the training of the Noise2Noise network.

In comparing performance with other algorithms, this experiment employs BM3D for conventional filter denoising and DnCNN for the supervised learning algorithm, which utilizes noisy and clean image training pairs. The primary focus is to compare performance with unsupervised learning image denoising algorithms, such as N2N, which generates noisy image pairs for training. S2S is implemented with a dropout rate of 0.2 in the learning structure during training. N2S sets random sampling to calculate the loss value at 20% of the training image size. SV-N2N introduces additive noise with a Gaussian noise parameter of $\sigma = 5$ to generate noisy image pairs in this experiment. For all unsupervised algorithms, a U-net structure is implemented, and training parameters are set similarly to the proposed self-augmented method.

VI. EXPERIMENT AND RESULT

In a system with high noise level interference, there may be a significant loss of important image features, leading to the possibility of overfitting in the proposed self-supervised learning method. Therefore, this research compares the learning outcomes of deep learning using only MSE loss with the combination of MSE and variance-based loss. This is because MSE alone may lead the network to perceive only average pixel-wise values during training, potentially overlooking the distribution of noise in the input. Consequently, this may result in a high pixel error when applying noise-augmented noise removal, as presented in the proposed method.

In Fig 6, (A) shows the result of using generated images from MSE learning for noise removal with the N2N framework. (B) shows the outcome of variance normalization with MSE learning applied to the noise removal process using the N2N framework. (C) shows the results of training over epochs, indicating that the loss measured by PSNR during training with and without variance normalization yields similar results, even though (B) addresses pixel errors more effectively.

The experimental comparison reveals that pixel errors are more prominent in flat regions, where the network often struggles to predict image features accurately. Despite achieving better PSNR in the validation set with just MSE loss during training, when applying the generated noisy images to the N2N denoising process, some residual noise remains. This suggests that relying solely on MSE loss may lead to more overfitting in the learning of the self-generated noisy image network compared to using MSE with variance normalization as a loss function. Normally, the learning of noise augmentation by the deep learning network can capture the noise distribution in the image through MSE loss. However, this experiment reveals that in training with 3-channel images, using MSE loss alone results in the generated noisy images, when used in the N2N network for denoising, retaining components of the noise. This is because the network learns only to approximate the average noise value of 3-channel images to be close to the input x_i'.

Additionally, Fig 7 shows the comparison of generated noisy images using various deep learning architectures. It can be observed that ViT can simulate noisy images closely resembling the initial noisy image, while ResNet and GANs exhibit abnormal pixel errors. This demonstrates that deep learning architectures are another factor influencing the generation of noisy images.

Table 2 and Fig 8 present the experimental results of noise elimination using generated Gaussian noise with $\sigma = 25$ and 50. The experiments illustrate that self-augmentation, employed to generate a noisy image dataset for Noise2Noise training, effectively reduces noise. Despite persisting pixel errors, these issues contribute to an overall enhancement in image denoising performance for the proposed method, as demonstrated in Table 2. Remarkably, the approach

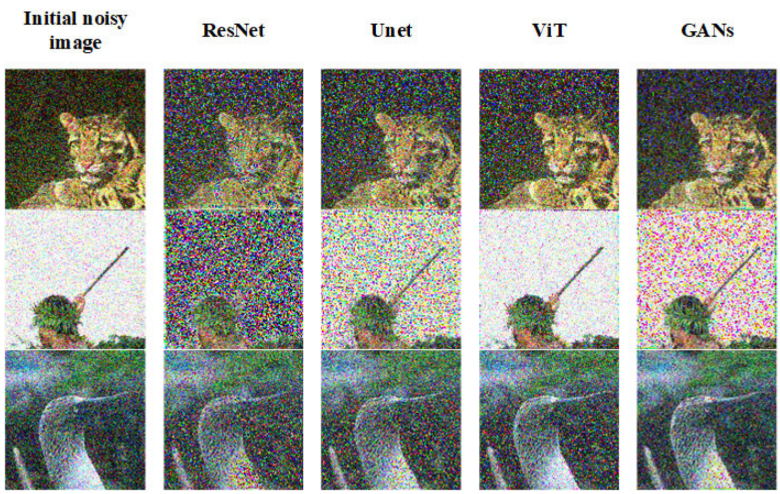


FIGURE 7. Example of noisy image augmentation results from each network model for Gaussian noise with a standard deviation of 50.

TABLE 2. Compare the PSNR (dB) and SSIM results between the proposed self-augmented noisy image for image denoising and other image denoising algorithms for the BSD300 dataset.

Algorithm	metric	BM3D	DnCNN	N2N (Unet)	N2S (Unet)	SV-N2N (Unet)	S2S (Unet)	Self-augmented (ResNet)	Self-augmented (Unet)	Self-augmented (ViT)	Self-augmented (GANs)
Gaussian	PSNR	23.59	30.20	29.50	24.69	25.68	25.34	21.84	25.53	25.40	21.15
$\sigma = 25$	SSIM	0.61	0.92	0.90	0.79	0.81	0.83	0.52	0.75	0.71	0.53
Gaussian	PSNR	19.06	26.83	25.92	20.19	22.26	24.47	18.40	20.91	23.08	15.70
$\sigma = 50$	SSIM	0.47	0.89	0.84	0.62	0.58	0.77	0.55	0.61	0.58	0.26

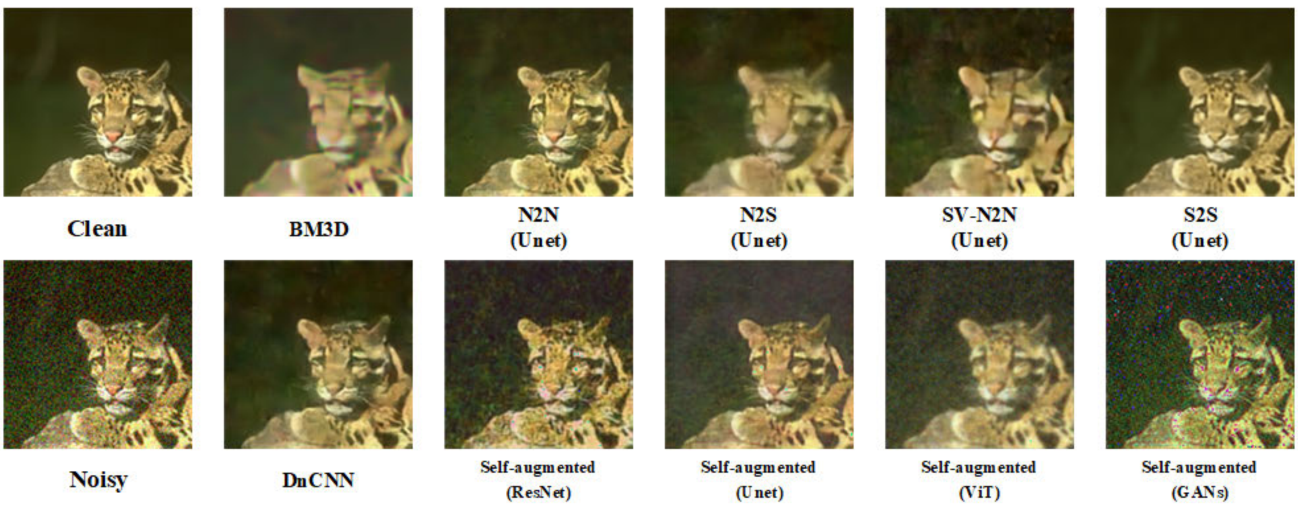


FIGURE 8. An example of image denoising of each algorithm for Gaussian noise with standard deviation (σ) of 25 in the BSD300 dataset.

achieves notable efficacy without the necessity of learning from a noise-free dataset.

Table 3 and Fig 9 provide a comparative analysis between self-augmentation and other algorithms for real-world image datasets. It is observed that the proposed method exhibits effective noise reduction, albeit less optimal than some alternative algorithms. As depicted in Fig 9, residual noise is still present in the images, suggesting potential limitations in generating noisy image datasets, particularly in scenarios with low noise levels when compared to the experiments in Table 2. This experiment highlights the problem of removing noise in images with low-level noise interference using the proposed method.

VII. DISCUSSION

From the experiment comparing the use of MSE with and without variance normalization in the U-net model, it is evident that normalization helps to align the distribution of

TABLE 3. Compare the PSNR (dB) and SSIM results between the proposed self-augmented noisy image for image denoising and other image denoising algorithms for real-world NIND and SIDD datasets.

Algorithm	metric	BM3D	DnCNN	N2N (Unet)	N2S (Unet)	SV-N2N (Unet)	S2S (Unet)	**Self-augmented (ResNet)**	**Self-augmented (Unet)**	**Self-augmented (ViT)**	**Self-augmented (GANs)**
NIND dataset	PSNR	27.78	25.36	23.83	23.64	23.96	24.91	20.79	20.65	20.31	23.61
	SSIM	0.68	0.71	0.61	0.52	0.46	0.43	0.32	0.32	0.31	0.46
SIDD dataset	PSNR	28.71	30.28	25.90	23.03	26.23	22.99	22.53	22.64	23.68	23.43
	SSIM	0.61	0.72	0.58	0.48	0.50	0.29	0.29	0.30	0.34	0.42

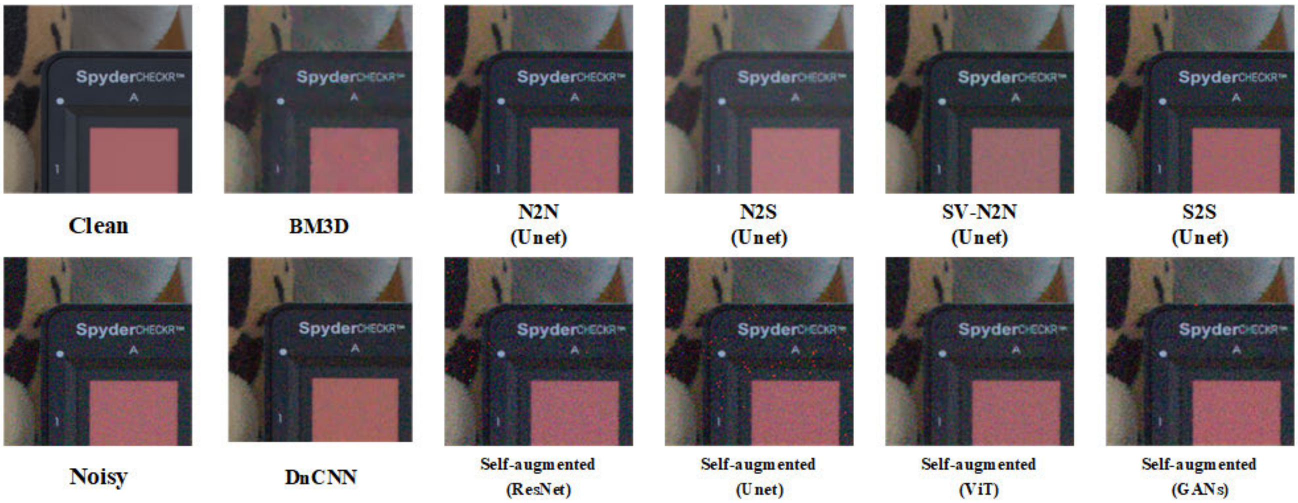


FIGURE 9. An example of image denoising of each algorithm for the NIND real-world noisy image.

generated noise pixels in the output image for each channel in the RGB color image system. This is because during deep learning training, the appropriate value of each pixel may not be considered, and the training process aims only to reduce the overall loss value across the dataset. Consequently, the generated noisy image output may assign excessive importance to certain channels, leading to the inability of Noise2Noise model to eliminate generated noise pixels in those channels when used for training. This inconsistency sometimes results in uncertain outcomes. Through this experiment, it is observed that variance normalization can help reduce the problem of noise pixel errors, but errors still occur occasionally, especially when the noise level is high, as shown in Fig 10.

For the comparison of the results of self-augmented learning using various deep learning structures, it is found that the skip-connected structure of ResNet tends to generate outputs with more pixel errors compared to other models. This might be due to feeding data from skip-connected layers leading to overfitting during learning noise augmentation. Conversely, the U-net model, being an autoencoder structure with the feature collapsing property in hidden layers, reduces noise features, resulting in fewer errors in the generated noisy image. However, since previous features are fed back through skip-connected layers, which undergo element-wise addition, it may lead to pixel errors during learning. As for the ViT model, which divides image patches and finds relationships between patches, it generates noisy images

FIGURE 10. The error in denoising Gaussian noise with σ=50 using self-augmented ResNet arises from abnormalities in the generated noisy image.

closest to the initial noisy image input. This is because self-identity learning enables ViT to utilize self-attention mechanisms to measure the relationship of noise pixels between patches, thus simulating output images with good alignment of noise pixels in each channel. In the case of GANs, where the generator competes with the discriminator, the discriminator often utilizes a feature extraction network to reduce input data size. This renders the generator unnecessary to simulate generated images similar to the initial noisy image to defeat the discriminator. This might cause the loss function, which is tied to the discriminator's outcomes, to halt the learning process of the generator, affecting the generated noisy image output.

In this experiment, the performance of the self-augmented strategy was compared to that of other image denoising algorithms. The findings reveal that the self-augmented method effectively generates noisy images for training the Noise2Noise framework, particularly in denoising Gaussian noise with a standard deviation of $\sigma = 25$ for U-net and ViT models. However, the ResNet and GANs models exhibited poorer noise removal results compared to other algorithms. When tested with Gaussian noise of $\sigma = 50$, the U-net model showed significantly degraded performance, suggesting potential overfitting issues possibly arising from the utilization of skip-connected layers. Additionally, when comparing the experimental results with other unsupervised image denoising methods, self-augmented using ViT models yielded slightly inferior results compared to other algorithms.

Considering the experimental results using real-world noisy images from the NIND and SIDD datasets, it was observed that the overall performance of the self-augmented strategy yielded inferior results compared to other unsupervised image denoising methods in terms of both PSNR and SSIM metrics in the NIND dataset. However, for the SIDD dataset, the results were relatively similar. This is likely because the self-augmented strategy in this experiment is based on the concept of zero-mean Gaussian noise, whereas real-world noisy images may not solely consist of Gaussian noise. As a result, the Noise2Noise denoising with generated noisy images may not effectively remove noise with different distributions, leading to residual noise in the denoised output.

VIII. CONCLUSION

This research focuses on developing the Noise2Noise algorithm to remove noise in single-image and blind noise scenarios by generating noisy images from limited noisy image datasets. In generating noisy images, the research introduces imperfections in the deep learning training process, which aims to optimize the training loss only for each training data, resulting in the output of the training process not perfectly mimicking the input. By using noisy images for training and validation, which constitutes self-identification learning, the deep learning model learns to generate new noisy images, which can then be used in subsequent Noise2Noise denoising processes. This research extends the use of self-augmentation image augmentation to single-image and blind denoising tasks. Experimental results show that self-augmentation in this research can remove noise in images similar to other unsupervised image denoising algorithms at low noise levels. However, there is significant degradation in noise removal performance at higher noise levels due to the proposed self-augmented noisy image network lacking understanding of the image features solely from perceiving noisy images. Furthermore, experimental results on real-world noisy image datasets show that the self-augmented strategy cannot effectively remove noise. This research demonstrates that image feature perception significantly impacts image restoration performance. Additionally, accurately estimating noise values and components in real-world noisy images is crucial for denoising algorithms to preserve important image features and minimize pixel errors, as demonstrated in the presented self-augmented approach.

REFERENCES

[1] R. A. Al Mudhafar and N. K. El Abbadi, "Noise in digital image processing: A review study," in *Proc. 3rd Inf. Technol. Enhance e-Learning Other Appl. (IT-ELA)*, Dec. 2022, pp. 79–84.

[2] N. Bindal, R. S. Ghumaan, P. J. S. Sohi, N. Sharma, H. Joshi, and B. Garg, "A systematic review of state-of-the-art noise removal techniques in digital images," *Multimedia Tools Appl.*, vol. 81, no. 22, pp. 31529–31552, Sep. 2022, doi: 10.1007/s11042-022-12847-7.

[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," *Inf. Fusion*, vol. 79, pp. 124–145, Mar. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521001792

[4] A. Wali, A. Naseer, M. Tamoor, and S. A. M. Gilani, "Recent progress in digital image restoration techniques: A review," *Digit. Signal Process.*, vol. 141, Sep. 2023, Art. no. 104187. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1051200423002828

[5] B. Goyal, A. Dogra, S. Agrawal, B. S. Sohi, and A. Sharma, "Image denoising review: From classical to state-of-the-art approaches," *Inf. Fusion*, vol. 55, pp. 220–244, Mar. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519301861

[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[7] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2808–2817.

[8] K. Bajaj, D. K. Singh, and M. A. Ansari, "Autoencoders based deep learner for image denoising," *Proc. Comput. Sci.*, vol. 171, pp. 1535–1541, Jan. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877050920311431

[9] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," 2018, *arXiv:1803.04189*.

[10] J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, M. Shahzad, W. Yang, R. Bamler, and X. X. Zhu, "A survey of uncertainty in deep neural networks," *Artif. Intell. Rev.*, vol. 56, no. S1, pp. 1513–1589, Oct. 2023, doi: 10.1007/s10462-023-10562-9.

[11] A. Shamsi, H. Asgharnezhad, A. Tajally, S. Nahavandi, and H. Leung, "An uncertainty-aware loss function for training neural networks with calibrated predictions," 2021, *arXiv:2110.03260*.

[12] M. Diwakar and M. Kumar, "A review on CT image noise and its denoising," *Biomed. Signal Process. Control*, vol. 42, pp. 73–88, Apr. 2018. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809418300107

[13] W. Feng, Y. Long, S. Wang, and Y. Quan, "A review of addressing class noise problems of remote sensing classification," *J. Syst. Eng. Electron.*, vol. 34, no. 1, pp. 36–46, Feb. 2023.

[14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jul. 2017, pp. 105–114, doi: 10.1109/CVPR.2017.19.

[15] A. Shocher, N. Cohen, and M. Irani, "Zero-shot super-resolution using deep internal learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, Jun. 2018, pp. 3118–3126. [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/CVPR.2018.00329

[16] Y. Liu, Z. Yan, J. Tan, and Y. Li, "Multi-purpose oriented single nighttime image haze removal based on unified variational retinex model," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 33, no. 4, pp. 1643–1657, Apr. 2023.

[17] Y. Liu, Z. Yan, S. Chen, T. Ye, W. Ren, and E. Chen, "NightHazeFormer: Single nighttime haze removal using prior query transformer," in *Proc. 31st ACM Int. Conf. Multimedia*, 2023, pp. 4119–4128. [Online]. Available: https://api.semanticscholar.org/CorpusID:258714816

[18] E. Chen, S. Chen, T. Ye, and Y. Liu, "Degradation-adaptive neural network for jointly single image dehazing and desnowing," *Frontiers Comput. Sci.*, vol. 18, no. 2, Jan. 2024, Art. no. 182707, doi: 10.1007/s11704-023-2764-y.

[19] C. Li, E. Hu, X. Zhang, H. Zhou, H. Xiong, and Y. Liu, "Visibility restoration for real-world hazy images via improved physical model and Gaussian total variation," *Frontiers Comput. Sci.*, vol. 18, no. 1, Jan. 2024, Art. no. 181708, doi: 10.1007/s11704-023-3394-0.

[20] 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," *Multimedia Tools Appl.*, vol. 79, nos. 27–28, pp. 20391–20427, Jul. 2020, doi: 10.1007/s11042-020-08815-8.

[21] S. Routray, A. K. Ray, and C. Mishra, "An efficient image denoising method based on principal component analysis with learned patch groups," *Signal, Image Video Process.*, vol. 13, no. 7, pp. 1405–1412, Oct. 2019, doi: 10.1007/s11760-019-01489-2.

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

[23] K. Bnou, S. Raghay, and A. Hakim, "A wavelet denoising approach based on unsupervised learning model," *EURASIP J. Adv. Signal Process.*, vol. 2020, no. 1, p. 36, Jul. 2020, doi: 10.1186/s13634-020-00693-4.

[24] J. Yoo and C. W. Ahn, "Image restoration by blind-Wiener filter," *IET Image Process.*, vol. 8, no. 12, pp. 815–823, Dec. 2014, doi: 10.1049/iet-ipr.2013.0693.

[25] D. N. H. Thanh, V. B. S. Prasath, L. M. Hieu, and S. Dvoenko, "An adaptive method for image restoration based on high-order total variation and inverse gradient," *Signal, Image Video Process.*, vol. 14, no. 6, pp. 1189–1197, Sep. 2020, doi: 10.1007/s11760-020-01657-9.

[26] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[27] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. 18th Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, vol. 9351, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[28] A. M. Ali, B. Benjdira, A. Koubaa, W. El-Shafai, Z. Khan, and W. Boulila, "Vision transformers in image restoration: A survey," *Sensors*, vol. 23, no. 5, p. 2385, Feb. 2023.

[29] T. Pang, H. Zheng, Y. Quan, and H. Ji, "Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2043–2052.

[30] A. Limsuebchuea, R. Duangsoithong, and J. Jaruenpunyasak, "A self-validation Noise2Noise training framework for image denoising," *Imag. Sci. J.*, pp. 1–16, Jul. 2023, doi: 10.1080/13682199.2023.2229040.

[31] A. Krull, T.-O. Buchholz, and F. Jug, "Noise2Void—Learning denoising from single noisy images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2124–2132.

[32] J. Batson and L. Royer, "Noise2Self: Blind denoising by self-supervision," in *Proc. Int. Conf. Mach. Learn.*, in Proceedings of Machine Learning Research, K. Chaudhuri and R. Salakhutdinov, Eds., Jun. 2019, pp. 524–533. [Online]. Available: https://proceedings.mlr.press/v97/batson19a.html

[33] Y. Quan, M. Chen, T. Pang, and H. Ji, "Self2Self with dropout: Learning self-supervised denoising from single image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 1887–1895.

[34] S. Guo, H. Yong, X. Zhang, J. Ma, and L. Zhang, "Spatial-frequency attention for image denoising," 2023, *arXiv:2302.13598*.

[35] A. Alsaiari, R. Rustagi, A. Alhakamy, M. M. Thomas, and A. G. Forbes, "Image denoising using a generative adversarial network," in *Proc. IEEE 2nd Int. Conf. Inf. Comput. Technol. (ICICT)*, Mar. 2019, pp. 126–132.

[36] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," *Sci. Rep.*, vol. 9, no. 1, p. 16884, Nov. 2019, doi: 10.1038/s41598-019-52737-x.

[37] F. Yue, C. Zhang, M. Yuan, C. Xu, and Y. Song, "Survey of image augmentation based on generative adversarial network," *J. Phys., Conf. Ser.*, vol. 2203, no. 1, Feb. 2022, Art. no. 012052, doi: 10.1088/1742-6596/2203/1/012052.

[38] H. Zhao, O. Gallo, I. Frosio, and J. Kautz, "Loss functions for image restoration with neural networks," *IEEE Trans. Comput. Imag.*, vol. 3, no. 1, pp. 47–57, Mar. 2017.

[39] A. Mustafa, A. Mikhailiuk, D. A. Iliescu, V. Babbar, and R. K. Mantiuk, "Training a task-specific image reconstruction loss," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2022, pp. 21–30.

[40] 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 IEEE Int. Conf. Comput. Vision. ICCV*, Jul. 2001, pp. 416–423.

[41] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Curves and Surfaces*, J.-D. Boissonnat, P. Chenin, A. Cohen, C. Gout, T. Lyche, M.-L. Mazure, and L. Schumaker, Eds. Berlin, Germany: Springer, 2012, pp. 711–730.

[42] A. Abdelhamed, S. Lin, and M. S. Brown, "A high-quality denoising dataset for smartphone cameras," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1692–1700.

[43] B. Brummer and C. De Vleeschouwer, "Natural image noise dataset," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1777–1784.

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, where he is currently pursuing the Ph.D. degree. His research interests include computer vision, deep learning, and machine learning.

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 the 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 Assistant Professor with the Department of Electrical and Biomedical Engineering, Faculty of Engineering, Prince of Songkla University. His current research interests include machine learning, computer vision, signal processing, and data analysis.

PORNCHAI PHUKPATTARANONT (Member, IEEE) received the B.Eng. (Hons.) 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 an Associate 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.