Article

Innovative Dual-Stage Blind Noise Reduction in Real-World Images Using Multi-Scale Convolutions and Dual Attention Mechanisms

Ziaur Rahman , Muhammad Aamir , Jameel Ahmed Bhutto , Zhihua Hu * and Yurong Guan *

School of Computer, Huanggang Normal University, Huanggang 438000, China;
ziaurrahman@hgnu.edu.cn (Z.R.); aamirshaikh86@hgnu.edu.cn (M.A.); jameel@hgnu.edu.cn (J.A.B.)
* Correspondence: huzhihua@hgnu.edu.cn (Z.H.); jsjgyr@hgnu.edu.cn (Y.G.)

Abstract: The distribution of real noise in images can disrupt the inherent symmetry present in many natural visuals, thus making its effective removal a paramount challenge. However, traditional denoising methods often require tedious manual parameter tuning, and a significant portion of deep learning-driven techniques have proven inadequate for real noise. Moreover, the efficiency of end-to-end algorithms in restoring symmetrical patterns in noisy images remains questionable. To harness the principles of symmetry for improved denoising, we introduce a dual deep learning model with a focus on preserving and leveraging symmetrical patterns in real images. Our methodology operates in two stages. In the first, we estimate the noise level using a four-layer neural network, thereby aiming to capture the underlying symmetrical structures of the original image. To enhance the extraction of symmetrical features and overall network performance, a dual attention mechanism is employed before the final convolutional layer. This innovative module adaptively assigns weights to features across different channels, thus emphasizing symmetry-preserving elements. The subsequent phase is devoted to non-blind denoising. It integrates the estimated noise level and the original image, thus targeting the challenge of denoising while preserving symmetrical patterns. Here, a multi-scale architecture is used, thereby amalgamating image features into two branches. The first branch taps into dilation convolution, thus amplifying the receptive field without introducing new parameters and making it particularly adept at capturing broad symmetrical structures. In contrast, the second branch employs a standard convolutional layer to focus on finer symmetrical details. By harnessing varied receptive fields, our method can recognize and restore image symmetries across different scales. Crucial skip connections are embedded within this multi-scale setup, thus ensuring that symmetrical image data is retained as the network deepens. Experimental evaluations, conducted on four benchmark training sets and 12 test datasets, juxtaposed with over 20 contemporary models based on the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) metrics, underscore our model's prowess in not only denoising but also in preserving and accentuating symmetrical elements, thereby setting a new gold standard in the field.

Keywords: deep learning; image denoising in real-world scenarios; attention mechanism; noise level assessment; multi-scale unit

1. Introduction

Within the domain of low-level vision applications, symmetry, as a fundamental aesthetic and structural principle, plays a pivotal role, and its preservation becomes especially crucial when it comes to image denoising. The meticulous restoration of symmetry is paramount in pre-processing, which is an essential step that leads to high-resolution image quality, accurate segmentation, precise detection, and flawless defect pinpointing. In real-world images, where symmetry is often a desired trait, noise can disrupt this inherent balance. Such disturbances predominantly arise from the innate constraints of digital devices, thereby compromising our ability to perceive and appreciate the symmetrical

essence of images. Historical examinations of image denoising literature indicate that early endeavors have been mainly focused on eradicating additive Gaussian white noise [1]. However, distinguishing this artificially induced noise from the asymmetry and distortions present in real-world images is crucial [2]. Despite technological advancements yielding cameras capable of capturing high-fidelity symmetrical visuals, challenges persist. External factors such as sensor limitations, uneven lighting conditions, and transmission errors can introduce asymmetry, and often, these disturbances are beyond control during the image acquisition process. The symmetry-disrupting complexity of noise only intensifies during the image transmission phase, thus getting influenced by the particularities of the transmission medium. Lastly, in-camera processing stages like demosaicing, grayscale adjustments, and data compression might either introduce new asymmetries or exacerbate existing ones [3]. Therefore, recognizing and restoring the balance and symmetry in images amidst these challenges is a cornerstone of contemporary image denoising endeavors.

In the process of confronting the complexities associated with image denoising, researchers have consistently emphasized the advancement of image denoising methodologies. Within this context, traditional denoising algorithms can be broadly categorized into two types: spatial domain filtering (SDF) [4] and frequency domain filtering (FDF) [5]. The SDF approach operates directly within the spatial domain of images by employing templates. In this method, the value of each pixel is determined by computing it based on the input pixel values through the template. On the other hand, FDF works by multiplying the image information within the frequency domain with specific algorithmic functions, thereby utilizing the Fourier transform [6] as its basis. Despite their utility, traditional denoising methods have inherent limitations. They can inadvertently omit crucial details, thus leading to outcomes that may not always meet the desired performance standards. Recognizing this limitation, researchers have explored alternatives. One such alternative involves the use of image prior models for denoising. A noteworthy contribution in this area is by Xu et al. [7], who introduced the non-local means (NLMs) denoising technique. This method capitalizes on the redundant information typically present in natural images. It identifies and averages similar regions within an image, using image blocks as the primary units, and subsequently removes the Gaussian noise present. Building on the concept of non-local denoising, Dabov [8] presented the BM3D algorithm. This innovative approach not only incorporates the principles of the non-local denoising method, but also synergizes it with wavelet transform domain denoising. As a result, it has demonstrated superior performance in various applications. However, a persistent challenge with these advanced methods is their reliance on pre-defined noise models. In the past, denoising algorithms typically used pre-defined noise models. These models are mathematical tools designed to represent common noise types in digital images, like Gaussian noise, salt-and-pepper noise, and speckle noise [9]. They work best when the noise's origin and characteristics are known, the noise remains uniform, and external factors introducing the noise can be predicted or controlled. This is often the case in lab settings or specific sectors where noise remains constant. However, as imaging technology changes, so does the nature of the noise. In real-world settings, especially outdoors, various unpredictable factors like changing lighting or movement can cause unique noise patterns. These can differ from the standard models, thereby making traditional denoising methods less effective. This highlights the need for techniques that can handle both established and unexpected noise types. Besides, it is evident that image denoising constraints manifest in various forms, which include, but are not limited to, total variation (TV) models [10,11], sparse models, and models incorporating self-similarity features. In particular, sparse representation methods have gained prominence for their ability to encapsulate the structural features of an image efficiently using an over-complete dictionary [12]. This approach has been especially fruitful in building effective variational models for denoising, thus yielding commendable results. Despite their flexibility and ease of interpretability, variational model-based algorithms are not without limitations. One significant challenge lies in the manual adjustment of parameters in the minimization process, which often necessitates expert knowledge. Furthermore,

these algorithms frequently employ prior constraints tailored to specific structural elements within images. This bespoke design approach compromises the generality of variational models, thus limiting their applicability across a broad range of image types. Elad et al. [1] emphasized that the effectiveness of a denoising algorithm is directly correlated with the performance of the subsequent tasks. Fundamentally, the objective of image denoising is to recover and reconstruct the inherent data within the image [13]. In the realm of image denoising, algorithms that leverage variational models have been a staple in traditional approaches. At the heart of these methods is the transformation of the image denoising challenge into an optimization problem governed by Bayesian principles, specifically a maximum posteriori probability formulation. This probabilistic problem is then recast as a minimization task, wherein the objective function is regularized using pre-established information about the image. The efficacy of these variational methods is profoundly influenced by the nature of the prior assumptions introduced into the model.

In recent times, the advancements in deep learning have significantly impacted various disciplines, including computer vision, pattern recognition, and image processing. These advances have also precipitated a marked improvement in the field of image denoising. Deep learning-based methodologies manifest superior adaptative capabilities, thus accommodating multi-faceted noise distributions while ensuring computational efficiency. The nascent phases of deep learning for image denoising witnessed the adoption of reinforcement learning strategies, such as policy gradients [14] and Q-learning [15], in the training of recursive neural architectures. However, the computational overhead and the inefficiency in optimization associated with these strategies were evident drawbacks. Modern deep learning denoising algorithms leverage architectural innovations such as skip connections, attention modules, and multi-scale feature integration to augment their feature representational capacity. Notwithstanding, the depth of these architectures occasionally culminates in training challenges, including gradient explosion or dissipation. A contemporary trend involves techniques like the AINDNet [16] and MPI_DA_CNN [17] adopting transfer learning paradigms [18] coupled with model compression strategies. This facilitates the transference of learned parameters from comprehensive models to more efficient architectures, thereby optimizing training processes and effectively circumventing gradient-related challenges. Furthermore, denoising strategies underpinned by graph neural networks (GNNs), exemplified by the GCDN [19] and GRDN [20], have underscored their merit, particularly in handling non-structured data, as well as capitalizing on the intrinsic topological attributes of graph networks.

Unlike their variational model-based counterparts, deep learning algorithms offer the advantage of being data-driven, thereby obviating the need for pre-defined prior assumptions. They are capable of autonomously learning relevant parameters from extensive datasets. However, a glaring drawback of these models is their lack of rigorous mathematical formulation, thus resulting in a deficit of interpretability. This poses a challenge in understanding the inner workings of these algorithms, which is crucial for their further refinement and application in diverse scenarios. The noise in real-world images is thus influenced by a confluence of both internal and external variables. This makes the noise profile particularly challenging to characterize and differs markedly from the noise that follows a specific distribution. In recent times, most deep learning-based denoising algorithms have been engineered to tackle additive Gaussian white noise, thus owing to its relative simplicity in both representation and removal. Training these algorithms often involves adding known levels of this specific noise to clean images, thus generating datasets comprising noisy and clean image pairs. However, such algorithms are specialized and tend to excel only at removing noise that follows the particular distribution [21]. This is a critical limitation, given that the distribution of noise in real-world images is rarely known a priori. Utilizing a denoising algorithm that assumes an incorrect noise level can have detrimental effects, which range from incomplete noise removal to the loss of important image details like edges [22].

To address these challenges, this paper introduces a novel neural network-based model specifically tailored for effective image denoising in complex, real-world scenarios. Our approach employs a two-stage network architecture that is optimized for both computational efficiency and denoising effectiveness. The model has been rigorously evaluated on commonly used image denoising datasets and has demonstrated robust performance in both quantitative metrics and visual assessment.

The main ideas and findings presented in this article are the following:

- We present a two-stage denoising network that is explicitly designed for denoising images captured in real-world scenarios. The first stage focuses on accurate noise level estimation, while the second stage performs targeted, non-blind denoising.
- We adopted a comprehensive approach that combined channel and spatial attention, thus leading to a new dual attention module. This module filters out less important information, thereby allowing essential data to be processed and conveyed with greater accuracy.
- Our design includes a specially crafted module that can extract features at multiple scales. This adaptability lets us tailor the network for various real-world scenarios by adjusting the quantities of these modules, thus balancing denoising efficiency with reduced network complexity.
- Based on our analysis of 16 benchmark datasets and using two metrics, when compared with over 20 traditional and contemporary algorithms, our proposed method proves to be both robust and adaptable. Furthermore, it quickly adjusts noise-degraded images and is suitable for a wide range of vision-based applications.

Through these contributions, this paper aims to advance the state of the art in image denoising, thus offering a robust, flexible, and efficient solution to an enduring challenge in low-level vision research.

2. Related Work

2.1. Traditional Denoising Algorithms

According to the characteristics of denoising methods, traditional denoising methods are divided into four categories:

Denoising using filtering: Commonly used techniques for image denoising include filtering methods [23,24]. Teng and Wang et al. [25,26] proposed an improved curvature filtering algorithm that uses a projection operator instead of the traditional curvature filter's minimum triangular cutting plane projection operator. They also modified the regular energy function to enhance the denoising ability. This algorithm has a good effect in removing strong noise but cannot adaptively adjust the projection operator in the neighborhood, and its operation time is longer. Moreover, Abazari et al. [27] developed a hybrid technique that combines the shearlet transform method with Yaroslavsky's filter for diverse image characteristics, including thin features and textures. Accordingly, the image is processed using the shearlet transform, followed by the application of the Yaroslavsky's filter, which is weighted based on pixel similarities from the previously denoised image. Goyal et al. [28] presented a computationally efficient algorithm that is based on non-local means combined with a non-subsampled shearlet transform (NSST). The source image is first decomposed using an NSST into coarser and finer layers. With two decomposition levels of the NSST, there is one set of low-frequency coefficients and four sets of high-frequency coefficients. The overall results are improved, but complex noise remains challenging to address. Liu et al. [29] completed image denoising by setting appropriate adjustment parameters, thereby dynamically selecting fixed thresholds and adding adjustment factors to reduce the constant deviation between the original wavelet coefficient and the estimated wavelet coefficient. Furthermore, You et al. [30] developed an image removal model that utilizes an enhanced wavelet transform combined with edge detection. This approach aids in increasing the image's signal-to-noise ratio while preserving as much edge information as possible. Al-Shamasneh et al. [31] introduced a method that uses local fractional entropy for estimating image pixel probability and employs quantum calculus to determine

the convolution window mask for image denoising. However, this model only removes Gaussian noise.

Denoising using sparse coding: Kumar et al. [32] introduced a model that employed weights to harmonize the varying scales of components within each group. These additional weights enhance the model's reconstruction accuracy and stability. However, the texture appears overly smoothed in the final result. Jia et al. [33] enhanced the BM3D algorithm by focusing on three areas: adaptive estimation of noise variance, domain transformation filtering, and non-linear filtering. While the model yielded improved visual outcomes, it resulted in excessively smooth edges. Mahdaoui et al. [34] presented a model using a compressed sensing reconstruction approach, thus the merging total variation regularization with the non-local self-similarity constraint. They optimized this method with an augmented Lagrangian, thereby sidestepping the challenges of non-linearity and non-differentiability inherent in the regularization terms. However, the trade-off is a minor increase in computational complexity relative to the image size, but this does not compromise real-time processing.

Denoising using external priors: Liu et al. [35] proposed a denoising algorithm that initially uses a shearlet to represent the input image sparsely. Then, it integrates non-local priors as constraints for image denoising with sparse representation. An alternating minimization algorithm is employed to solve this constrained denoising issue, thus yielding the denoised image. However, this algorithm has not been tested on other types of coherent images. Bhargava et al. [36] used singular value decomposition and hard thresholding methods to collaboratively denoise the obtained multi-scale similar matrix. Qi et al. [37] introduced image structural priors and sparse priors into image restoration processing by proposing two improved algorithms based on TV and sparse representation. Xie et al. [38] proposed a deep learning-based approach for reducing noise in images. This method employs an optimization function that includes non-local regularizers. These regularizers have two parts: a spatial filter and a frequency domain filter. Their purpose is to encourage the sparsity of gradients in the solution. The symmetric U-Net achieved better results.

Denoising using low-rank representation: Fan et al. [39] imposed a TV norm constraint on the coefficient matrix in the low-rank representation model, thereby proposing a novel image denoising method. Luo et al. [40] integrated relative total variation (RTV) into a weighted nuclear norm minimization (WNNM), thus imposing an RTV norm constraint on the WNNM low-rank representation model. However, the construction of the image denoising model and the optimization process have high computational costs, thereby leading to a longer processing time. Buades [41] introduced an algorithm that uses the redundant information commonly found in natural images. Unlike commonly used bilinear filtering and median filtering that utilize local image information for filtering, it uses the entire image for denoising. It searches for similar areas in the image based on image blocks, averages these areas, and effectively removes Gaussian noise present in the image.

Regarding traditional denoisers, they often use only noisy images for training and denoising. Many effective denoising algorithms are based on the BM3D algorithm. This algorithm's idea is somewhat similar to NL-Means. The BM3D algorithm is currently the most effective traditional image denoising method. However, due to the special and complex nature of image noise, there are few similar blocks in complex texture areas that are mostly edge areas, thereby resulting in suboptimal denoising effects and leading to the loss of detail, blurring, and other phenomena.

2.2. Advances in Deep Learning-Based Denoising Algorithms

Based on the type of noise present in the image, deep learning denoising algorithms can be categorized into four distinct classes:

Denoising Additive Gaussian White Noise Images: Zhang et al. [42] introduced a convolutional neural network (CNN) model that integrates batch normalization with residual learning techniques. While this method yielded notable results, the algorithm necessitates extensive iterations to secure an optimal training model, thus compromising

its efficiency and convergence rate. Valsesia et al. [19] developed a method utilizing graph convolution operations to establish a non-local receptive field. This model, termed graph–convolutional image denoising (GCDN), leverages dynamic similarity calculations within hidden features. Nonetheless, the proposed architecture remains unextended to other inverse problems, such as super-resolution. Subsequently, Wang et al. [43] presented a streamlined image denoising network featuring an innovative four-channel interaction transformation. They adjusted both the input and output images to include four channels, thus adding an extra channel filled with zeros. This added channel, set to approach zero during training, guides the training procedure and boosts the network's resilience to errors.

Denoising Real-World Noise Images: Yan et al. [44] innovatively extracted noise patterns directly from degraded images, thereby achieving unsupervised noise modeling and successfully denoising unpaired real noise images. Their framework is constructed on the self-consistent generative adversarial network (SCGAN) paradigm. Meanwhile, Zhao et al. [45] offered a solution for dark burst images by employing a recurrent fully convolutional network (RFCN), thus directly mapping raw burst imagery to standard red–green–blue (sRGB) outputs. While this approach boasts considerable flexibility, it has not been adapted for video denoising, and its portability remains a challenge.

Denoising Blind Noise Images: Yang et al. [46] proposed a pioneering strategy using a multi-column CNN for estimating the noise level function from singular images. This technique, however, has yet to be applied to natural image denoising. Yu et al. [47] introduced a deep iterative down-up CNN capable of cyclically adjusting the resolution of feature maps, thereby allowing it to manage various noise intensities using a singular model without the need for supplementary noise information. Chen et al. [48] presented the GAN–CNN-based blind denoiser, which exploits the GAN for noise distribution modeling, thus generating noise samples and collaborating with clean image datasets to train a denoising network. A limitation to consider is its presupposition of zero-mean additive noise. Moreover, Bian et al. [49] introduced a denoising algorithm comprising four components: sparse representation, initial feature fusion, attention mechanism, and residual module. The sparse representation component extracts local features from the image, while the feature fusion component merges both global and local features, thereby augmenting the network's ability to represent the image.

Denoising Mixed Noise Images: Zhang [50] devised a tri-layered super-resolution network furnished with a dimensional augmentation strategy, thus creating a versatile framework competent in addressing numerous or spatially varying degradations. Shah et al. [51] presented a two-stage model that is based on patch transformation specifically for mixed noise elimination. They combined this with a bilateral filtering method to preserve image edges. This was integrated into a cognitive neural network model aimed at denoising images. The network's inherent adaptability detects the presence of mixed noise and creates a training dataset consisting of both noisy and denoised patches. An area warranting further investigation is the trade-off between the network complexity and performance. Several existing methods employ deep networks with numerous layers, thus inevitably leading to protracted training periods. Conversely, shallow networks may compromise denoising performance. Recent innovations, therefore, focus on developing two-stage real image denoisers that utilize channel attention mechanisms for improved noise level estimation, along with multi-scale modules for non-blind denoising, which aim to reconcile effective denoising with computational efficiency.

3. Theoretical Foundation

To thoroughly comprehend the algorithmic architecture proposed in this study, it is imperative to anchor our methodology within recognized theoretical tenets. Denoising algorithms, particularly those leveraging CNNs, are underpinned by the conventional convolution operation and CNNs' inherent capability to hierarchically discern feature representations from datasets. Owing to their architectural design, CNNs intrinsically discern patterns across multiple scales and complexities, thereby rendering them especially suited

for endeavors such as image denoising. The preliminary phases of our model, as depicted in Figure 1, are premised on the understanding that noise in authentic images predominantly emerges as stochastic perturbations in pixel intensities. Traditionally, scholars have characterized this noise as Gaussian, given its ubiquity and amenable mathematical attributes. The decision to split the denoising process into two stages, noise measurement and targeted denoising, comes from the understanding that noise in real-world settings is not consistent, thus requiring accurate assessment for effective removal. The dual attention module, illustrated in Figure 2, is rooted in the paradigm of 'attention' in neural processing [52]. The human cerebrum does not equitably process every piece of incoming information; rather, it prioritizes specific stimuli contingent upon context, antecedent experiences, or intrinsic relevance. In parallel, attention mechanisms within neural networks are designed to differentially weight input features, thereby empowering the network to accentuate pertinent attributes, which are paramount for the given task. Dilated convolutions [53], as showcased in Figure 3, stem from the objective of amplifying the receptive purview of convolutional procedures. Within standard convolution, each resultant element is contingent upon a limited proximate region in the input. By interspersing gaps, dilated convolutions allow each output component to consider an expansive input domain, thus effectively augmenting the contextual view without escalating computational demands.

Conclusively, the paradigm of residual learning, exemplified in Figure 4, is predicated on the assertion that discerning residual mappings, i.e., disparities or deviations, is frequently more straightforward and efficacious than direct mappings. Residual pathways, by facilitating direct interconnections across layers, alleviate the vanishing gradient dilemma, thus capacitating the training of more profound networks with efficiency. With the theoretical underpinnings of our methodology articulated, the ensuing sections delineate the specific algorithmic enhancements and their practical realizations.
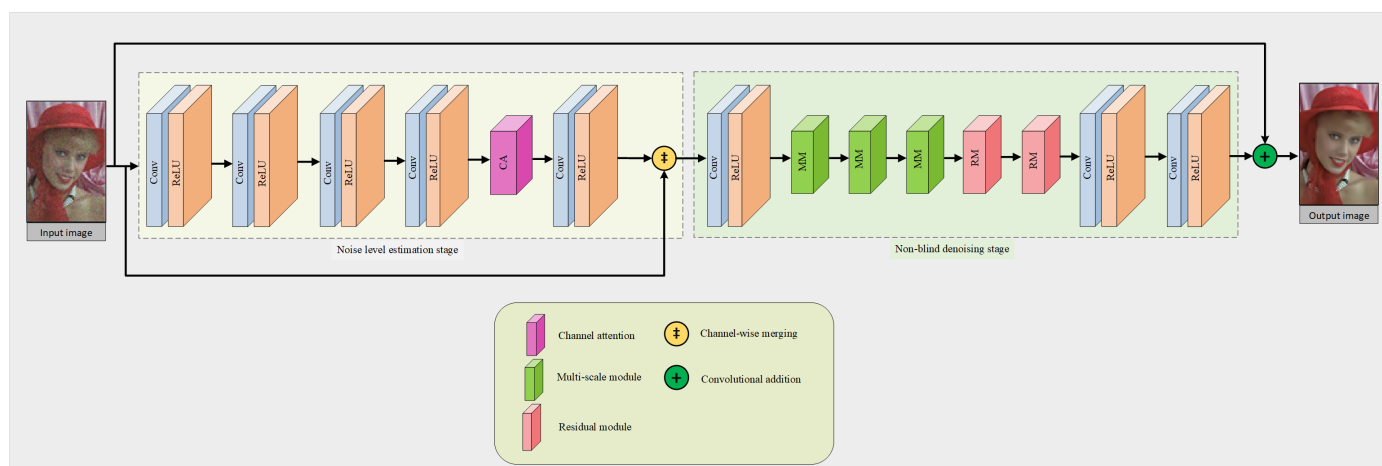

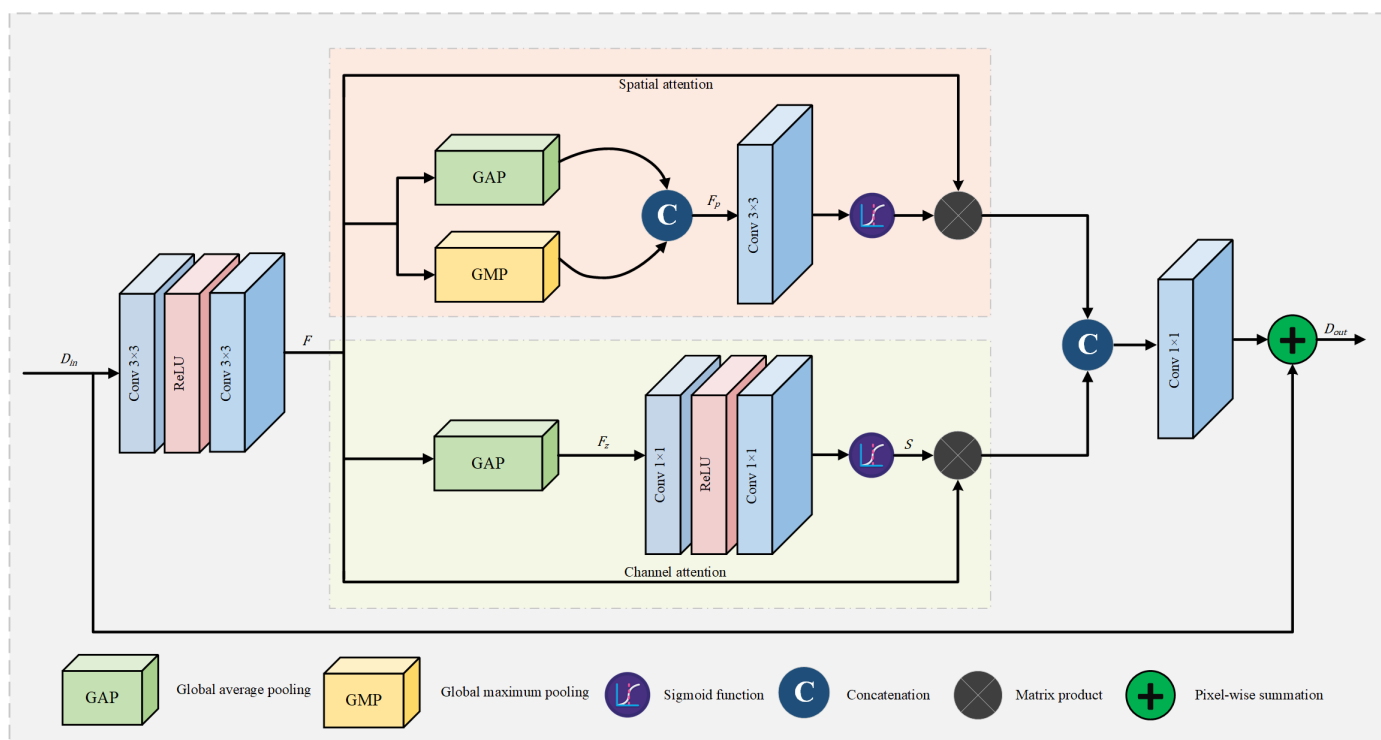**Figure 1.** Overall structure of the proposed model.
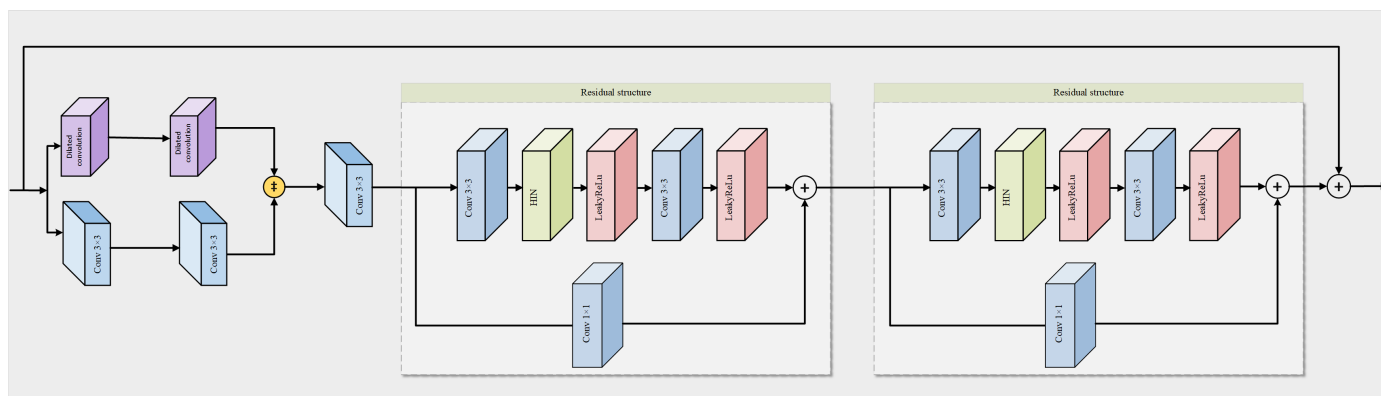
Figure 2. Representation of the dual attention module.

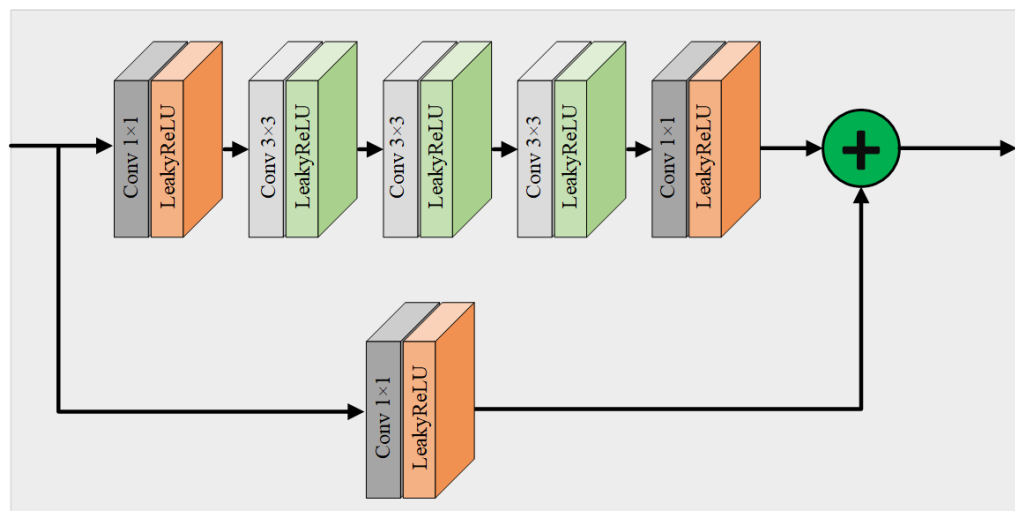**Figure 3.** Graphical representation of mulit-scale structure.

Figure 4. Graphical representation of residual structure.

4. Algorithmic Framework

Denoising algorithms specifically engineered for mitigating noise within predetermined distribution levels, such as the DnCNN [42] and FFDNet [54], commonly adopt an end-to-end computational paradigm to directly attenuate noise artifacts in image data. However, it should be emphasized that the distribution level of noise in real-world image scenarios is inherently uncertain. This intrinsic ambiguity poses considerable challenges to the effectiveness of a solitary end-to-end algorithmic process in successfully denoising authentic images.

To circumvent this limitation, an initial step that focuses on the quantitative estimation of the noise distribution level present in the real-world images is indispensable. Subsequently, this quantitatively estimated noise level is combined with the original image and fed into the subsequent stages of the algorithmic process. By doing so, the complex problem of denoising real-world images can be reduced to the more tractable issue of eradicating Gaussian noise at a specified distribution level. It is important to highlight that existing denoising algorithms have shown significant effectiveness in reducing noise within certain distribution parameters. This proven performance provides essential prior knowledge that can be used to fine-tune the denoising method discussed in this paper. Therefore, the denoising approach outlined in this research is divided into two main stages. The first stage focuses on estimating the noise level in the noisy images. Following this is the second stage, which carries out the targeted denoising of the images. A visual diagram of this approach is available in Figure 1.

4.1. Dataset Pre-Processing

In deep learning-based techniques, dataset pre-processing is essential for enhancing the reliability and uniformity of the input data, which in turn guarantees that the models are trained using dependable and characteristic examples. Due to the intrinsic high resolution of the images in the SIDD dataset, there is a necessity for image pre-processing to facilitate efficient model training. To this end, each original image was segmented into multiple non-overlapping patches of size (256×256) pixels. Our decision to adopt the (256×256) resolution is based on initial experiments, which assessed the model's efficacy over a range of resolutions spanning from 128×128 to 512×512. The selected size strikes an optimal balance between computational demands and model precision. This resolution aptly represents a considerable number of use cases within our intended application domain. To accommodate real-world images of varying sizes, we employ a pre-processing strategy to adjust them to the chosen resolution. This size has been selected after preliminary experiments, which showed a balanced trade-off between retaining essential details and computational feasibility. However, to validate the performance and robustness of the proposed model for image denoising, we employed a diverse set of benchmark datasets specifically designed for this purpose. These datasets cover a wide range of noise types, intensities, and real-world scenarios, thus ensuring a comprehensive evaluation of the model's capabilities. A detailed description of the characteristics and origins of these datasets is provided in the following subsections that illuminate their relevance and significance within the context of our study.

4.1.1. Training Datasets

The training dataset [13] is bifurcated into two primary categories: grayscale noise-affected images and their color counterparts. The dataset for grayscale-noise-impacted images is designed for the training of both Gaussian and blind denoising algorithms. Within this classification, two notable datasets are present: the BSD400 and the Waterloo Exploration. The former, BSD400, encompasses 400 images stored in the .png format. These images have been resized to 200×200 pixels to suit the requirements of denoising model training. On the other hand, the Waterloo Exploration dataset is comprised of 4744 authentic images, which are also in the .png format. As for color-noise-affected images, they are represented in the BSD432, Waterloo Exploration, and polyU datasets. Delving

into specifics, the polyU dataset houses 100 genuine noisy images of 2784×1856 pixel dimensions, which have been sourced from five different camera models. These cameras include the Nikon D800, Canon 5D Mark II, Sony A7 II, Canon 80D, and the Canon 600D.

4.1.2. Test Datasets

As delineated in reference [13], our evaluation datasets span a variety of image collections, which are marked by both grayscale and color noise. The grayscale-noise-affected image collection amalgamates three distinct datasets: Set5, Set12 and BSD68. Specifically, Set12 incorporates 12 distinct scene images, whereas BSD68 aggregates a compilation of 68 natural images. These collections are instrumental in gauging the efficacy of Gaussian denoising techniques and blind noise attenuation algorithms. Pertaining to the color image dataset, we assimilated an eclectic mix of datasets, namely: CBSD68, Kodak24, Urban100, McMaster, CC15, DND, NC12, SIDD, and CC60. Of these, the Kodak24 and McMaster datasets comprise 24 and 18 color-noise-affected images, respectively. The DND dataset is distinguished by its 50 authentic noisy images, with their pristine counterparts sourced from minimal ISO settings. Conversely, the NC12 dataset houses 12 noisy images, but it notably lacks their clean equivalents. The SIDD dataset is particularly noteworthy; it encapsulates genuine noisy images acquired via smartphones, thus amounting to 320 paired sets of noisy images and their pristine counterpart images. Concluding our collection, the Nam dataset integrates 11 unique scenes, which are all archived in the JPEG format.

4.2. Quality Metrics

The assessment of image quality depends on utilizing quality metrics, which offer objective evaluations of their performance. These metrics are essential for evaluating the efficacy of denoising methods and for comparing various techniques. By measuring the amount of noise reduction and the resemblance between the denoised and the original, noise-free image, these metrics offer important information about the visual quality and fidelity of the denoised images. In this paper, the quality of the denoised images is evaluated using the PSNR and SSIM metrics, which are shown in Equations (1) and (2), respectively [55]. In general, higher values of the PSNR and SSIM indicate improved visual quality of the enhanced results.

$$PSNR_{dB} = 20 \times \log_{10} \left(\frac{MAX[N(i,j)]}{\frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} [N(i,j) - K(i,j)]^2} \right) \quad (1)$$

$$SSIM = \frac{(2u_1 u_2 + c_1)(2\sigma_{1,2} + c_2)}{\left(u_1^2 + u_2^2 + c_1\right)\left(\sigma_1^2 + \sigma_2^2 + c_2\right)} \quad (2)$$

where $N(i,j)$ represents the pixel value at position (i,j) of the original noise-free image, and $K(i,j)$ is the pixel value at position (i,j) of the denoised image. H and W represent the height and width of the image, respectively. u_1 and u_2 are the means of $N(i,j)$ and $K(i,j)$, respectively. σ_1 and σ_2 represent the variances of $N(i,j)$ and $K(i,j)$, respectively. $\sigma_{1,2}$ indicates the covariance between $N(i,j)$ and $K(i,j)$. The constants $c_1 = 0.01$ and $c_2 = 0.02$ are introduced to maintain system stability.

4.3. Phase of Noise Level Estimation in Image Data

The principal objective of this particular phase is to rigorously estimate the level of noise that is inherently present in real-world images. A mathematical model to describe real-world noisy images can be formalized as $(y = x + n(x))$, where y symbolizes the image corrupted by noise, x stands for the original, uncorrupted image, and $n(x)$ refers to the actual noise modeled as a Gaussian distribution $N(0, \sigma^2(x))$. It is noteworthy to mention that in the subsequent non-blind denoising phase of the neural network, the target is to eliminate the noise, represented by $n(x)$, from the noisy images. However, the challenge lies in the unknown status of the parameter σ that describes the level of noise distribution,

which subsequently hinders the network's ability to establish a precise mapping from noisy to clean images. Therefore, the results emanating from this initial noise level estimation phase are imperative; they serve as the crucial distribution level parameter σ that feeds into the non-blind denoising stage.

To address this complex issue, this study proposes a specialized sub-network solely dedicated to the task of noise level estimation. This segment employs a fully connected CNN, consisting of four layers, to make estimations regarding the noise level in the given input image. Importantly, this section forgoes the use of pooling layers and does not engage in batch normalization procedures, thus maintaining the complexity and feature richness of the original data. Instead, it opts for convolutional layers with a kernel size of 3×3 and systematically conducts padding operations subsequent to each convolutional layer. This is to ensure the conservation of the feature dimensions across different layers. Moreover, recognizing the proven efficacy of attention mechanisms in the extraction of salient features and the consequent enhancement of network performance, we added a dual attention mechanism; the details are given in the subsequent sections.

4.3.1. Dual Attention Mechanism

Neural networks traditionally handle spatial and channel features indiscriminately, which can inhibit their capability to recognize and prioritize essential features. Such an indiscriminate approach could potentially curtail the potential depth and richness of the networks' representations. Given that not all features significantly influence the denoising capacity of a network, it is paramount for such networks to allocate more computational focus to pivotal features. Integrating attention mechanisms has been identified as an effective approach to address this concern. Originating from the domain of machine translation [56], attention mechanisms have been seamlessly embedded into contemporary neural architectures, thus finding widespread applications across natural language processing, statistical learning, and computer vision disciplines. A helpful analogy is the human visual system: our eyes tend to concentrate on significant segments of visual input, thus sidelining the less pertinent regions.

To accentuate the importance of this, during the intricate processes of encoding and decoding, emphasizing key features can substantially elevate the denoising efficacy of the network. Recent breakthroughs, especially the infusion of attention mechanisms in primary vision tasks [57], have provided the motivation for this study. We embarked on a holistic approach that amalgamated channel attention (CA) [58] and spatial attention (SA) [59], thereby resulting in a novel dual attention module. This integrated module is meticulously designed to inhibit less critical information, thus paving the way for vital data to be processed and transmitted with enhanced fidelity, as depicted in Figure 2. The governing principle behind this is articulated as follows:

$$D_{out} = D_{in} + Conv1([SA(F) * CA(F)]) \tag{3}$$

where the feature map represented as $F \in R^{H \times W \times C}$ is derived from the dual attention module, post the dual convolution process on the input tensor $D_{in} \in R^{H \times W \times C}$. The term $Conv1$ signifies a convolution layer characterized by a 1×1 kernel size.

4.3.2. Channel Attention (CA)

The essence of CA can be conceptualized as dynamically weighting each channel. A channel with a pronounced weight underscores its heightened relevance to pivotal information. As feature maps scale in dimensionality, the spatial dimensions shrink, yet the number of channels proliferates. This surge can challenge the neural network's proficiency in distinguishing salient channel information. Nevertheless, adopting the CA mechanism can spotlight channels that are of paramount importance, thereby often yielding commendable outcomes.

The CA branch of our module operates on the principle of squeezing and exciting inter-channel correlations in convolutional feature mappings. The initial step involves

spatially compressing features, which is succeeded by an excitation phase, which astutely captures the intricacies of inter-channel dynamics [60]. Utilizing the global average pooling (GAP) operation on the feature map F effectuates this compression, thus morphing the $H \times W \times C$ feature dimensions into $F_z \in R^{1 \times 1 \times C}$. The simplicity of the global average pooling operation ensures a universal receptive field, thereby allowing even the lower echelons of the network to harness global insights. Subsequent to this, the excitation process calibrates F_z via dual convolution layers, followed by a sigmoid activation function to yield normalized weights in the 0 to 1 range, thus culminating in weights $S \in R^{1 \times 1 \times C}$. Conclusively, these weights undergo multiplication with the original feature map F to recalibrate the features in the channel dimension, thus producing the output of the CA branch.

4.3.3. Spatial Attention (SA)

In CNNs, each layer's output typically conforms to the dimensions $H \times W \times C$, with C signifying channels, and H and W representing the dimensionally reduced height and width, respectively. SA addresses all channels in a two-dimensional spatial configuration, thus engendering a weight matrix for a feature map of the dimensions $H \times W$. Each pixel assimilates a distinct weight, thus symbolizing the prominence of its spatial location. Appending this weight matrix to the prototypical feature map augments valuable features and diminishes the less significant ones [61].

In line with this, the SA branch seeks to recalibrate the input feature map F based on its spatial interrelationships, thus formulating a dedicated spatial attention map. To generate this map, both the global average pooling and global maximum pooling (GMP) are applied to the feature map F across its channel dimension. The resulting outputs are concatenated to produce the feature map $F_P \in R^{H \times W \times 2}$, which is then funneled through a convolution layer and a sigmoid activation function. This generates the spatial attention map that then recalibrates the original feature map F, thus leading to the final output of the spatial attention branch.

4.4. Non-Blind Denoising Stage

4.4.1. Multi-Scale Denoising Module Architecture

Within the framework of the non-blind denoising stage, a nuanced multi-scale architecture is developed. This architecture is dichotomized into two distinct branches. The first branch is characterized by two convolutional layers, with each employing a kernel size of 3×3. Concurrently, the second branch also incorporates two convolutional layers with identical kernel dimensions of 3×3. However, what demarcates the second branch from the first is the deliberate utilization of a dilated convolutional strategy.

Dilated convolution is instrumental in augmenting the receptive field by interpolating gaps into the feature map. This approach has the potential to capture more comprehensive and large-scale contextual information from the input data. In the context of our design, both layers in the second branch exhibit a dilation rate of two and adhere to specific padding strategies to ensure the size consistency of the resulting feature maps across layers.

Subsequent to this, the feature maps derived from these parallel branches are subject to fusion. This composite feature map is then combined with residual connections, thus serving as the input for succeeding layers in the neural network architecture. This architecture is ingenious in its ability to harness the extensive feature information gathered through dilated convolution while simultaneously preserving the integrity of local features. The amalgamation of these divergent types of feature information culminates in a feature set that is robust and comprehensive.

The structural design of this multi-scale module is delineated in Figure 3. In this schematic representation, layers rendered in lighter hues indicate the employment of dilated convolution. The symbol "$(\oplus)$" is employed to signify the element-wise addition of corresponding channels in the feature map, thereby emphasizing the integration aspect of this architectural design.

4.4.2. Residual Structure

The role of preserving intricate details in images is paramount, particularly in problems requiring restoration processes such as image denoising. Conventional approaches utilizing batch normalization for dimensionality reduction within neural networks have been found to compromise pixel-level information [62]. While normalization techniques have become ubiquitous in a plethora of computer vision tasks, the indiscriminate application of these methods does not necessarily yield favorable outcomes. Specifically, empirical evidence suggests that batch normalization fails to contribute to performance enhancement in super-resolution applications [63].

In addressing these limitations, our research deploys specialized residual blocks for feature extraction, the architecture of which is depicted in Figure 5. Post-processing via these residual blocks enables the propagation of feature maps to subsequent computational modules in the network. Contrary to conventional residual blocks, our modified architecture deliberately omits the batch normalization layer. Instead, we introduced a half-instance normalization (HIN) strategy. This innovative approach bifurcates the feature maps along the channel dimension: the first sub-set of channels undergoes normalization via the instance normalization (IN) technique, whereas the second sub-set is designed to retain contextual information intrinsic to the image [64]. The residual structure can be seen in Figure 4.
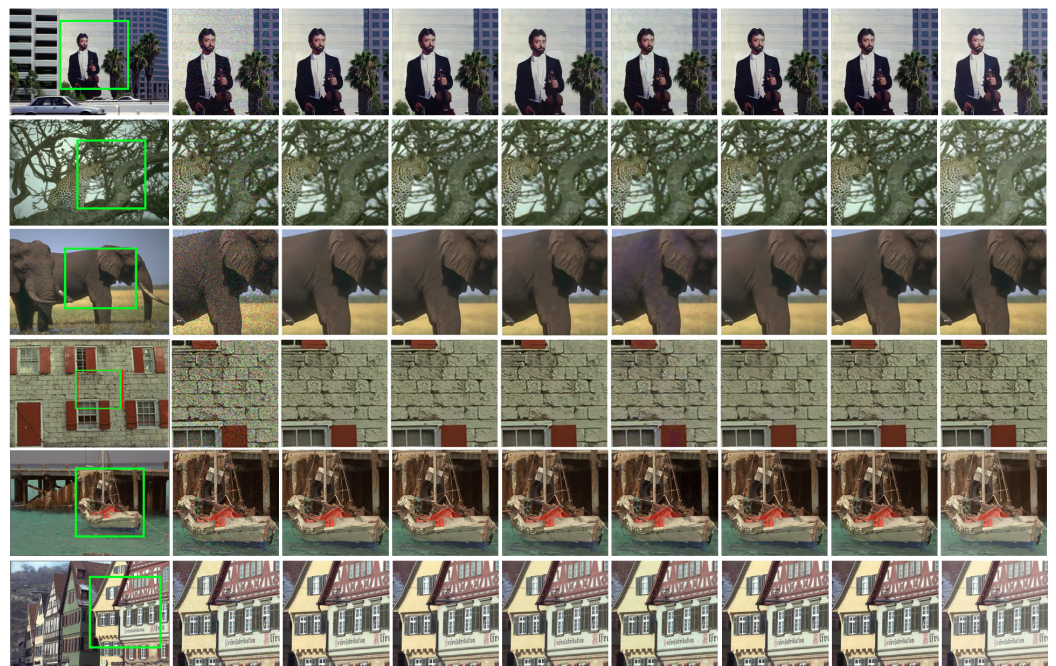

Figure 5. Comparison of visual results between previous methods and our proposed approach. The images in the first three rows are sourced from the BSD68, while the images in the last three rows are sourced from the Kodak24.

The duality of this configuration serves a dual purpose; it allows for the normalization of essential features while preserving contextual cues that are vital for image restoration tasks. This union of normalized and non-normalized features is achieved through channel-wise concatenation. Furthermore, we employ leaky rectified linear units (Leaky ReLUs) as the activation function within our residual blocks, thereby setting the negative slope parameter to 0.2. The architecture culminates in the utilization of 3×3 convolutional layers, which are designed to compute the residual outputs essential for image reconstruction. This novel approach not only safeguards the retention of critical image details, but also paves the way for enhanced restoration capabilities.

4.4.3. Loss Function Selection: A Case for L1 Loss

In the context of image denoising, $\ell1$ and $\ell2$ loss functions are ubiquitously utilized. The $\ell1$ loss function computes the aggregate absolute deviations between the true and predicted pixel values in an image. Conversely, the $\ell2$ loss function, which squares the errors, tends to amplify any discrepancies between the true and predicted pixel values, thus leading to a heightened penalty for larger errors. Given that the $\ell2$ function is prone to produce artifacts such as blurring and loss of detailed features, we chose to employ the $\ell1$ loss function for model optimization. Let I_{real} represent the real image and I_{pred} represent the predicted or denoised image. Both I_{real} and I_{pred} have the same dimensions: height H and width W. The pixel values at location (i, j) in the images are given by $I_{real}(i, j)$ and $I_{pred}(i, j)$, respectively. The mathematical representation of our chosen loss function is the following:

$$L1(I_{real}, I_{pred}) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} |I_{real}(i, j) - I_{pred}(i, j)| \tag{4}$$

In conclusion, the architecture and loss function employed in our model have been meticulously chosen to optimize performance, especially under challenging conditions.

5. Experimental Results

5.1. Implementation Details

Model Training: Our neural network underwent a rigorous training process, which was characterized by the following:

- Training Duration: The model was trained extensively over a span of 4000 epochs. This duration was determined based on the convergence behavior observed during preliminary runs.
- Learning Rate Adaptation: An initial learning rate of (10^{-4}) was set for the first 1500 epochs. Post this, to ensure finer weight updates and to stabilize convergence, the learning rate was decayed by a factor of one-tenth every subsequent 1000 epochs.

Computational Details: The training and other computational tasks were orchestrated using the following setup:

- Hardware: We relied on the robust NVIDIA GTX 2080ti GPU, which ensured efficient parallel processing and reduced training times.
- Software Framework: All neural network components, including layers, optimizers, and loss functions, were implemented using the PyTorch framework, which provided flexibility and ease of experimentation.

5.2. Qualitative and Quantitative Assessment

To assess the denoising performance of traditional techniques in comparison with deep neural network-based methods, we executed both quantitative and qualitative evaluations on a wide range of datasets. Our quantitative assessment employed pivotal metrics like the PSNR and SSIM, thereby offering a numerical evaluation of the denoised image quality. Concurrently, our qualitative evaluation utilized visual representations to demonstrate the restored images, thereby providing an intuitive sense of their visual quality and accuracy. This holistic assessment method ensured a thorough insight into the denoising capabilities of different techniques over various datasets, thus marking a significant contribution to image processing research.

5.2.1. Denoising Color Images

In this subsection, we performed experiments to evaluate the efficacy of our proposed model, thus contrasting it with existing models. We initially tested our model's proficiency on color images utilizing two datasets: the BSD68 and Kodak24. To maintain an equitable comparison, our evaluation encompassed six blind denoising techniques: the AVMF [65], DeGAN [66], SFAA [67], NNF [68], FCNN [69], and DIBS [70].

Furthermore, we pitted it against three non-blind denoising methods: the BM3D [71], TWSC [72], and FFDNet [54]. Visual comparisons, as depicted in Figure 5, elucidate our findings. Notably, when tested on the SIDD dataset, the FFDNet, TWSC, BM3D, and DeGAN fell short of effectively counteracting real noise. Conversely, while the DIBS and FCNN demonstrated noise mitigation capabilities, they tended to compromise on image texture nuances and edge details. In juxtaposition, our method showcased superior visual results, thus marking it as a notable advancement.

In a second phase of our experiment, we further evaluated the performance of the proposed model by applying it to color images from the SIDD dataset. Figure 6 provides insightful visual comparisons, thus highlighting the challenges faced by several established denoising methods when tackling the SIDD dataset, such as the FFDNet, TWSC, BM3D, and DeGAN, which struggled to effectively eliminate genuine noise. It is worth noting that while the DIBS and FCNN were successful in noise reduction, they tended to sacrifice essential image texture details and edge information. In contrast, our approach consistently demonstrated superior visual results, thus making it a promising solution for noise reduction in diverse images.
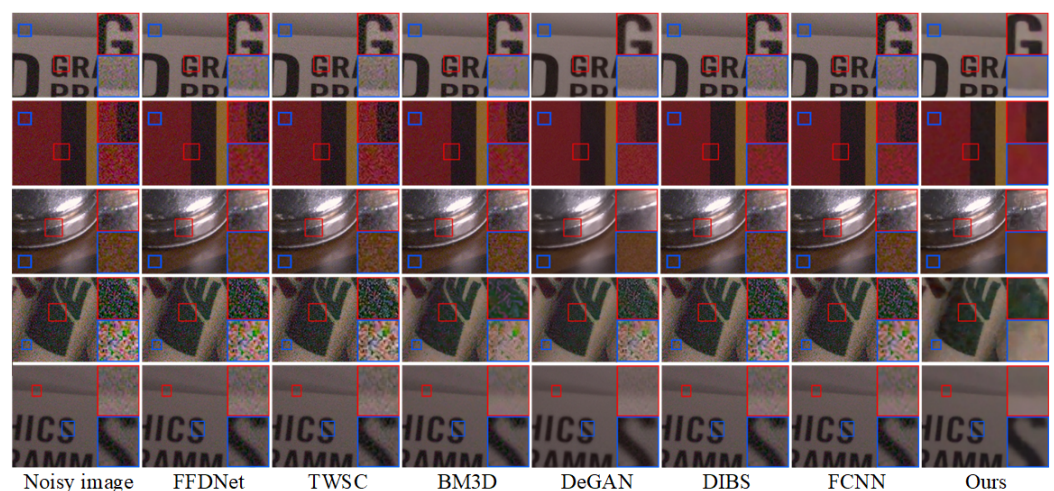

Figure 6. Denoising results of prior and proposed approaches on SIDD dataset.

As a third step in our evaluation, we extended our assessment of the proposed model's performance to color images obtained from the DND dataset. In Figure 7, the results clearly illustrate that our method excelled in maintaining image details without introducing unwanted artifacts. In contrast, the other methods tended to sacrifice the integrity of the edge structures and finer texture details, thus highlighting the effectiveness of our approach for preserving image quality and detail in images.
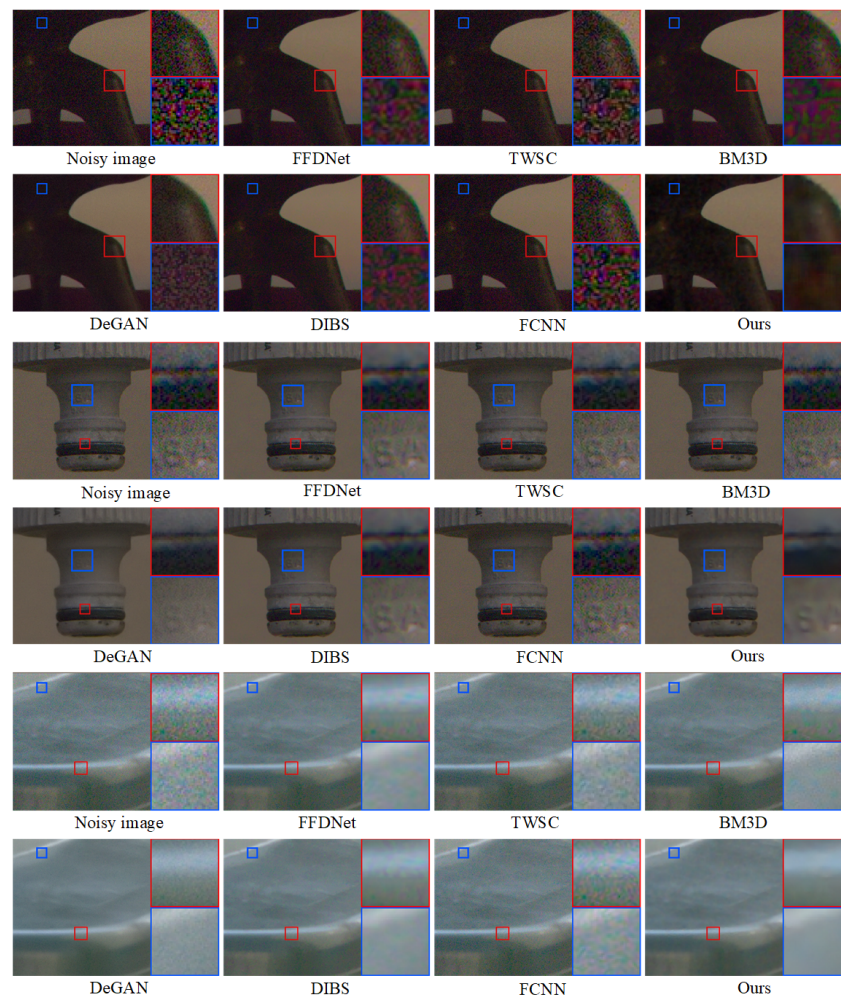
Figure 7. Denoising results of prior and proposed approaches on the DND dataset.

As a continuation of our evaluation, we extended our analysis to encompass color images sourced from the RNI15 dataset. Figure 8 presents compelling evidence of our method's superior performance. It not only exceled in noise reduction, but also effectively mitigated artifacts, preserved essential edge information, and produced images that are not only clearer, but also more visually appealing. In addition, the other prior models such as the S2S-LSD [73], FCNN [69], DnCNN [42], and NNF [68] also obtained quality results. However, the proposed model's findings underscore the versatile benefits of our proposed methodology in enhancing image quality and minimizing undesirable elements in diverse datasets, thereby establishing its credibility as a valuable tool for image enhancement and restoration.

Figure 8. Denoising outcomes of previous and proposed approaches on the RNI15 dataset.

Upon quantitative evaluation of the proposed model across three distinct datasets, insightful patterns in denoising performance emerge. Table 1 showcases the PSNR values of the various methodologies for different noise levels. For the CBSD68 dataset, our method consistently outperformed all other techniques across the tested noise levels, thereby achieving a PSNR of 34.89, 31.79, and 28.89 for noise levels of 15, 25, and 50, respectively. This superiority was mirrored in the Kodak24 and McMaster datasets. Particularly in the Kodak24 dataset, our model surpassed the other approaches, with a notable difference at a noise level of 50, where our method attained a PSNR of 29.86. Similarly, in the McMaster dataset, our approach continued its dominance by achieving the highest PSNR values for all the noise levels: 35.88, 33.23, and 30.01, respectively. These findings underline the effectiveness of our proposed technique in comparison to other state-of-the-art methods, such as the BM3D, DnCNN, DIBS, FFDNet, S2S-LSD, and NNF.

In addition, we employed a paired t test to compare the PSNR values of our method with those of other methods at each noise level. A paired t test is a statistical test commonly used to compare the means of two paired samples. It is particularly suitable when the paired samples are not independent, such as when different methods are applied to denoise the same set of images. We selected the paired t test for this analysis due to its robustness and insensitivity to violations of the assumption of normality. The results of the paired t tests are presented in Table 1. The calculated p values on the CBSD68 dataset demonstrate that our approach significantly outperformed the BM3D, DnCNN, DIBS, S2S-LSD, and NNF methods in terms of performance. However, there was no significant difference in performance between our method and the FFDNet method. Furthermore, our method exhibited a significant improvement over the BM3D method on the Kodak24 dataset. However, there was no significant difference in performance between our method and the DnCNN, DIBS, FFDNet, S2S-LSD, and NNF methods. Finally, when analyzing the McMaster dataset, our method performed similarly to the BM3D and DIBS methods. However, the comparison between our method and the DnCNN, FFDNet, S2S-LSD, and NNF methods did not yield conclusive results.

Table 1. Quantitative evaluation of various methods based on PSNR across three datasets.

Datasets	Noise level	BM3D [71]	DnCNN [42]	DIBS [70]	FFDNet [54]	S2S-LSD [73]	NNF [68]	Ours
CBSD68	15	33.42	33.81	33.79	33.91	34.11	33.88	**34.89**
	25	30.71	31.11	31.17	31.19	31.39	31.29	**31.79**
	50	27.41	27.89	27.79	27.96	28.09	28.04	**28.89**
	p value	0.009475	0.01719	0.027771	0.019703	0.036744	0.034747	
Kodak24	15	34.32	34.6	34.69	34.63	34.88	34.76	**35.31**
	25	32.18	32.09	32.19	22.11	32.51	32.28	**32.89**
	50	28.51	28.89	28.91	29.01	29.31	29.45	**29.86**
	p value	0.031629	0.008397	0.016818	0.344042	0.012156	0.012579	
McMaster	15	34.06	33.39	34.58	34.77	35.28	35.22	**35.88**
	25	32.66	31.52	32.21	32.45	32.85	32.66	**33.23**
	50	28.62	28.72	28.82	29.28	29.62	29.42	**30.01**
	p value	0.075244	0.034982	0.004811	0.018127	0.023792	0.023792	

5.2.2. Denoising Grayscale Images

In the domain of grayscale image processing, particularly during the denoising phase, we embarked on a rigorous evaluation of the proposed model using images from three renowned datasets: Set12, Set14, and BSD68. The visual outcomes of these models, presented in Figure 9, have been derived after training them with a plethora of noisy images at noise intensities of 25 and 50. Our analytical lens is further broadened by juxtaposing our method against established approaches like the S2S-LSD [73] and DnCNN [42]. Distinctive trends can be observed: the S2S-LSD [73] often led to overly smoothed outputs, thus sacrificing edge clarity and texture details, whereas the DnCNN [42] was prone to generating darker, more blurred denoised images. Contrarily, our proposed technique stood out by proficiently preserving sharp contours and rich details while simultaneously ensuring aesthetic coherence in smoother areas of the images.

Furthermore, we collected various traditional and learning-based approaches to validate their performance outcomes against the proposed model. Figure 10 displays the average PSNR results of the different methods on the Set5, Set12, and NC12 datasets with noise levels of 15, 25, and 50. When $\sigma = 15$, the BM3D [71], EPLL [74], TNRD [75], and DnCNN [42] attained lower PSNR scores compared to the ADNet [76], FOCNet [77], GCDN [19], NNF [68], and S2S-LSD [73]. However, our method attained a score of 33.25, which was the highest among the evaluated methods. It was closely followed by the S2S-LSD [73] with a score of 33.16 and the GCDN [19] at 33.14. For $\sigma = 25$, our method remained in the lead, scoring 30.94, whereas the S2S-LSD [73] and GCDN [19] were close competitors with scores of 30.81 and 30.78, respectively. At $\sigma = 50$, our approach continued to hold the leading position with a score of 27.82, while the FOCNet [77] scored 27.68 and the S2S-LSD [73] scored 27.64. The consistent superiority of our method implies enhanced generalization ability across various noise levels, as indicated by the different sigma values, in contrast to the other examined techniques.
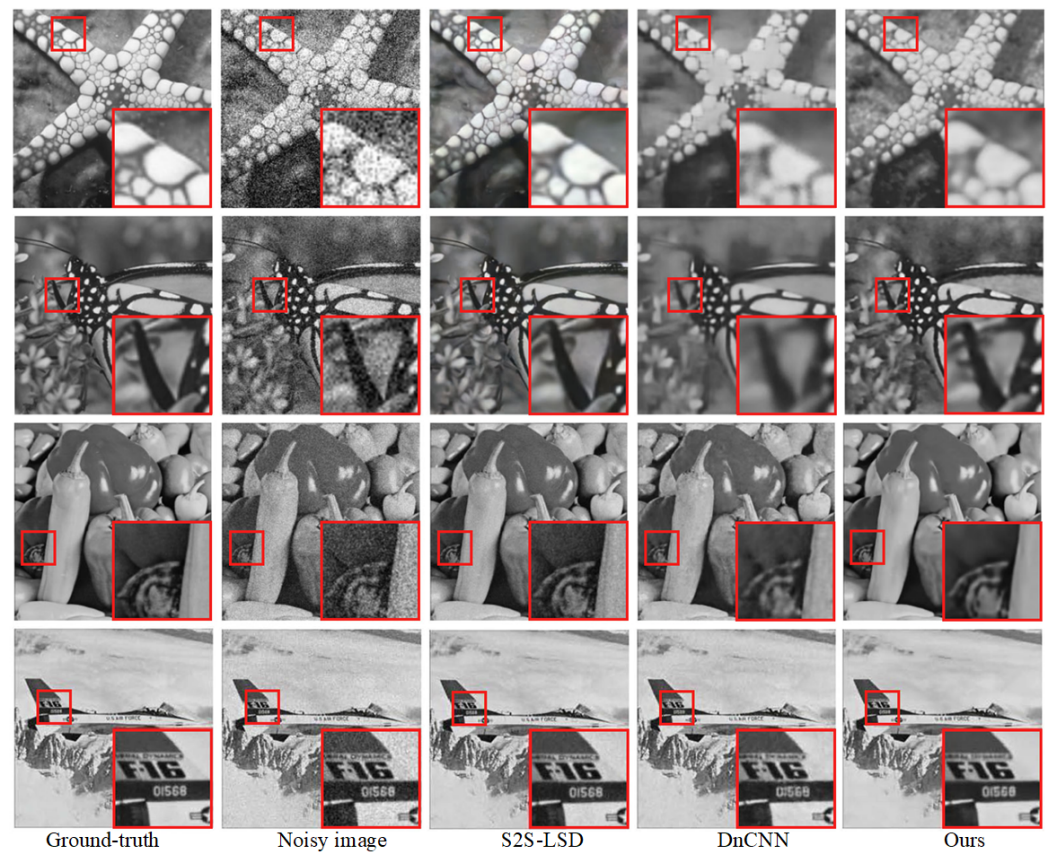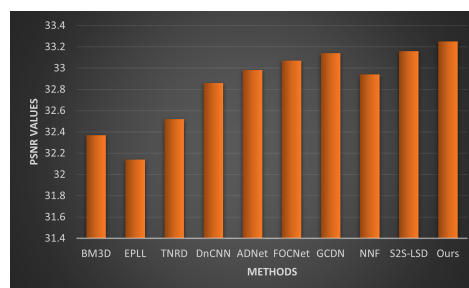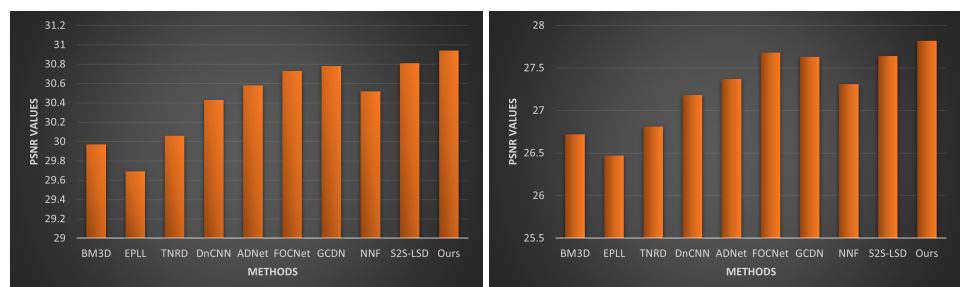
Figure 9. Visual outcomes of grayscale images.

(**a**) Denoising results with noise level σ = 15

(**b**) Denoising results with noise level σ = 25 (**c**) Denoising results with noise level σ = 30

Figure 10. Quantitative assessment of proposed and prior approaches.

Deep learning techniques for denoising images with mixed types of noise present a substantial challenge in real-world scenarios. This is because images with damage often have a variety of noise, thereby complicating the restoration of the original, clean image. To combat this issue, the idea of multi-degradation leveraging deep learning

technology is introduced. This article presents the performance of such a multi-degradation model for image denoising. In Table 2, we provide a comparative assessment of various denoising methods based on the PSNR and SSIM metrics, with a particular focus on the Bicubic downsampling degradation [78]. The results, taken from four datasets, i.e., Set5, Set14, BSD100, and Urban100, illustrate that our method exhibited competitive or superior performance across different scales when compared to other state-of-the-art techniques.

We conducted a statistical analysis using paired t tests to compare our method with previous approaches. In order to determine the superior method, we examined the p values of each method in relation to a pre-determined significance level, typically set at 0.05. If the p value is less than 0.05, we reject the null hypothesis and conclude that there is a statistically significant difference in performance between our method and the method being compared. Conversely, if the p value is greater than 0.05, we fail to reject the null hypothesis and conclude that there is insufficient evidence to suggest a significant difference in performance. Based on the SSIM values and the measured p values obtained from the Set5 dataset, we are unable to conclude that our method is significantly superior or inferior to the NNF, FCNN, DnCNN, and S2S-LSD methods. The p values indicate that there is no significant difference in performance between our method and the other methods. However, when considering the provided p values from the Set14 dataset, our method performed significantly better than the NNF method. Nevertheless, there was no significant difference in performance between our method and the FCNN, DnCNN, and S2S-LSD methods. Similarly, when examining the p values from the Urban100 dataset, our method outperformed the NNF method significantly. However, there was no significant difference in performance between our method and the FCNN, DnCNN, and S2S-LSD methods. Finally, with regard to the BSD100 dataset, our method performed significantly better than the NNF and FCNN methods. However, there was no significant difference in performance between our method and the DnCNN and S2S-LSD methods.

When considering the PSNR values and the measured p values, our method demonstrated a significant improvement in performance compared to the NNF, FCNN, and DnCNN on the Set5 dataset. However, there was no significant difference in performance between our method and the S2S-LSD. Similarly, on the Set14 dataset, our method outperformed the NNF, FCNN, and DnCNN significantly, but there was no significant difference in performance between our method and the S2S-LSD. On the Urban100 dataset, we cannot definitively determine whether our method is significantly better or worse than the NNF, FCNN, and S2S-LSD. However, our method did exhibit a statistically significant difference in performance compared to the DnCNN. Lastly, on the BSD100 dataset, our method demonstrated a significant improvement over the NNF and FCNN. However, there was no significant difference in performance between our method and the DnCNN or S2S-LSD.

In addition, Figure 11 showcases the PSNR and SSIM metrics for the various denoising methods applied to the CC15, CC60, and Nam datasets. It is noteworthy that the TWSC and BM3D often held their own against deep neural network-based methods like the CycleISP [79], DnCNN [42], FFDNet [54], NNF [68], DANet [80], and S2S-LSD [73] across these datasets. The performance of learning-based techniques is not always superior to that of their traditional counterparts, which may be due to constraints in the training data. The complexity of real-world noise means it does not always follow expected patterns. This unpredictability has spurred the development of blind denoising techniques, especially those rooted in deep learning. Evaluating the efficacy of previous methods, including the DnCNN [42], FFDNet [54], NNF [68], and DANet [80], yielded commendable results. Yet, in contrast to these learning-based techniques, our approach demonstrated superior denoising performance.

Table 2. Quantitative comparison of various denoising methods using PSNR and SSIM metrics across Set5, Set14, BSD100, and Urban100 datasets under different scale factors.

Datasets	Scale Factor	NNF [68]		FCNN [69]		DnCNN [42]		S2S-LSD [73]		Ours	
		SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR
Set5	2	0.929	33.64	0.953	36.62	0.959	37.58	0.959	37.66	**0.969**	**37.79**
	3	0.868	30.39	0.908	32.74	0.922	33.75	0.923	**33.93**	**0.925**	32.45
	4	0.81	28.42	0.863	30.48	0.885	31.4	0.886	31.58	**0.893**	**31.96**
p value		0.040788	0.034573	0.043145291	0.286164818	0.078202309	0.786163382	0.113155947	0.634808932		
Set14	2	0.868	30.22	0.906	32.42	0.913	33.03	0.913	33.19	**0.915**	**33.32**
	3	0.774	27.53	0.821	29.27	0.832	29.81	0.834	29.94	**0.837**	**30.04**
	4	0.702	25.99	0.751	27.48	0.767	28.04	0.77	28.18	**0.777**	**28.35**
p value		0.016861	0.00715	0.074865608	0.002147566	0.135841435	0.007463862	0.12011731	0.022352567		
Urban 100	2	0.841	26.66	0.897	29.53	0.914	30.74	0.916	31.02	**0.92**	**31.33**
	3	0.737	24.46	0.801	26.25	0.828	27.15	0.833	27.38	**0.84**	**27.57**
	4	0.657	23.14	0.722	24.52	0.752	25.2	0.758	25.35	**0.773**	**25.68**
p value		0.011694	0.032585	0.043368568	0.017686966	0.096437539	0.009895742	0.118517445	0.02406957		
BSD100	2	0.844	29.55	0.887	31.34	0.775	31.9	0.897	32.01	**0.898**	**32.05**
	3	0.738	27.2	0.786	28.4	0.798	28.85	0.799	28.91	**0.803**	**28.97**
	4	0.667	25.96	0.71	26.9	0.725	27.2	0.726	27.35	**0.734**	**27.49**
p value		0.004222	0.022013	0.043897929	0.004882648	0.359176086	0.070531418	0.166049611	0.12011731		

(**a**) PSNR values on CC15

(**b**) PSNR values on CC60

(**c**) PSNR values on Nam

(**d**) SSIM values on CC15

(**e**) SSIM values on CC60

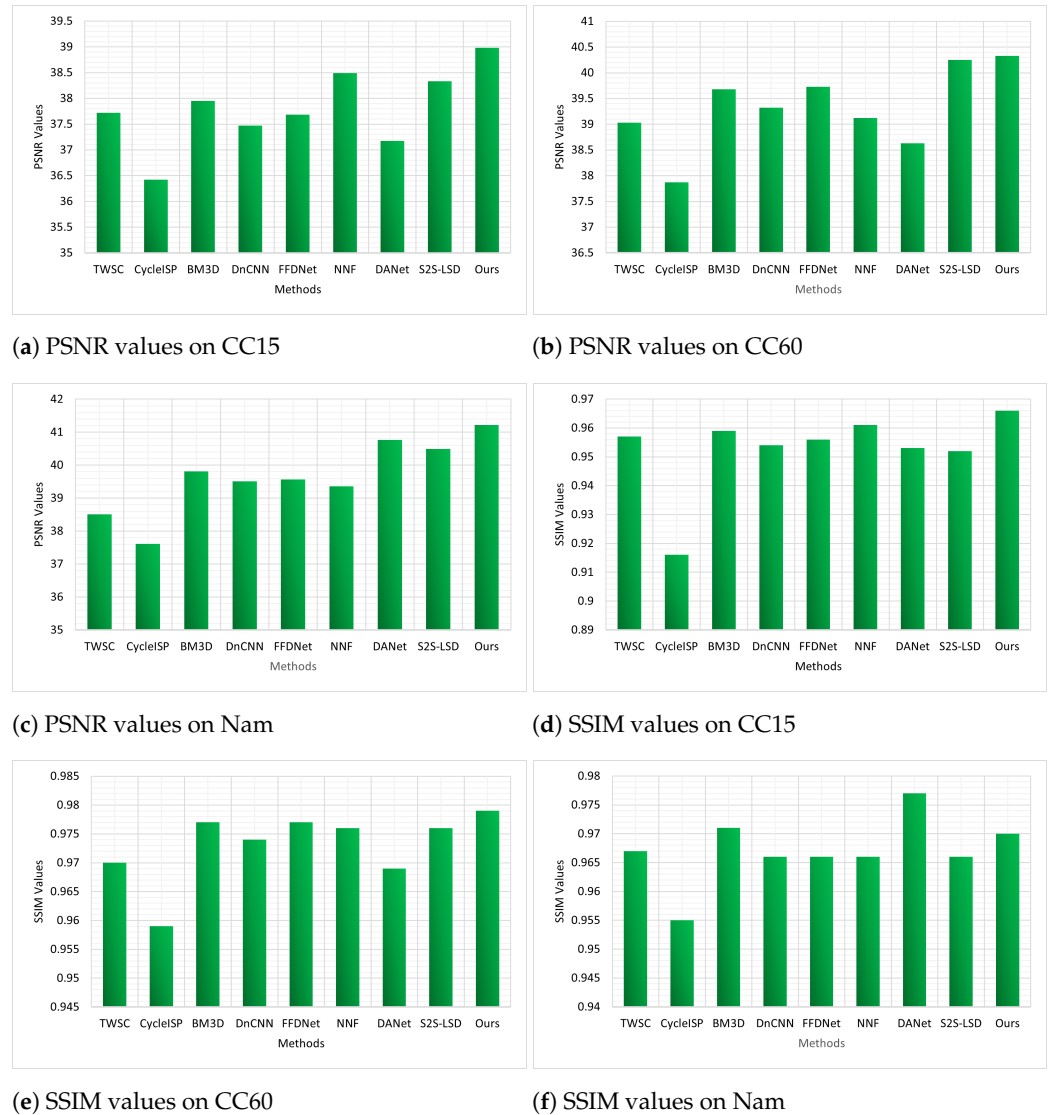(**f**) SSIM values on Nam

Figure 11. Comparative performance outcomes of PSNR and SSIM metrics for traditional and learning-based methods on the CC15, CC60, and Nam datasets.

5.3. Computational Complexity

We selected certain deep learning algorithms for a comparative study with our proposed algorithm. For the network configuration, we used an input image of size 256 × 256 with three channels. Table 3 details the parameter count and computational complexity for the entire network. Regarding execution time, we utilized a consistent environment to denoise images of the same size. Our experiments show that, when benchmarked against the DnCNN [42], ADNet [76], and NNF [68], our algorithm took less time. The increase in running time can be attributed to the prior algorithm's deeper network structure, which inherently adds to its parameters and computational intricacy. Nonetheless, our method demonstrates superior denoising capabilities compared to the alternatives.

Table 3. Comparative analysis of parameters and execution times across various denoising algorithms.

Method	Running Time/s (Per Image)	Time to Train the Model	FLOPs (10^9)	Parameter (10^6)
DnCNN [42]	0.98	-	1.46	0.55
ADNet [76]	0.86	-	1.36	0.52
NNF [68]	0.7	-	1.27	0.48
Ours	0.21	15 h	1.11	0.91

6. Discussion

Implications of the Results: Our newly developed deep learning network, designed specifically for denoising real-world images, brings a new approach to image restoration. The method divides the process into two stages: initially estimating the noise level and subsequently focusing on targeted denoising. This systematic approach is crucial for handling image denoising, especially when aware of the noise distribution, which is a common issue in computer vision and image processing.

Strengths and Limitations: Our approach stands out due to the dual attention mechanism and the multi-scale structure, both of which are key in recognizing various image features across different scales. The method of using dilation convolution to expand the receptive field without adding extra parameters deserves mention. However, while our model introduces several innovations, it does not surpass some of the highly advanced models in the field. These advanced models can identify a wider range of image features that our model may overlook.

Potential Applications: Our denoising model is versatile and can be applied to several computer vision applications, such as image segmentation, object detection, and facial recognition, wherein image clarity is crucial. Additionally, in areas like image compression and enhancement, the effectiveness of our denoising technique proves valuable.

Future Directions: Even though our model demonstrates promising results, there is room for improvement in future iterations. The aim would be to enhance the model to ensure it captures vital image data while keeping the architecture simple and adaptable. Exploring ways to merge our model's strengths with the extensive feature recognition of advanced networks holds promise for future research.

Concluding Remarks

This research introduces a novel deep learning network tailored for denoising real-world images. Our approach adopts a two-stage process for image restoration. In the initial stage, we determine the noise level using a neural network comprised of four layers. To bolster feature extraction and elevate the network's efficacy, we incorporate a dual attention mechanism module preceding the ultimate convolutional layer. This module dynamically allocates weights across various feature channels. The subsequent stage zeroes in on non-blind denoising, thus leveraging both the deduced noise level and the image. This addresses the unique challenge of denoising when the noise distribution is known. We also introduce a multi-scale framework, thus fusing image features across twin branches. One branch employs dilation convolution, thereby augmenting the receptive field without introducing additional parameters, whereas the other adopts a conventional convolutional layer. This strategy ensures the capture of diverse image traits across multiple receptive fields. To mitigate information loss with increasing network depth, strategic skip connections are embedded within the multi-scale framework. Through rigorous testing on four benchmark training sets, 12 test datasets, and by juxtaposing our model with over 20 established counterparts using PSNR and SSIM metrics, our technique sets a new performance standard.

Although our two-stage blind denoising network is effectively and thoughtfully crafted for image denoising, some intricate, sophisticated, learning-based models [81–83] trained on complex noisy images might outperform simpler learning-based models. These

advanced networks have high computational costs and are adept at recognizing a wider range of image features. As a result, a crucial area for future research is to refine our network's architecture, thus balancing the retention of essential image information while maintaining a design that is both simple and adaptive.

References

1. Elad, M.; Kawar, B.; Vaksman, G. Image denoising: The deep learning revolution and beyond—A survey paper. *SIAM J. Imaging Sci.* **2023**, *16*, 1594–1654. [CrossRef]
2. Zhou, L.; Zhou, D.; Yang, H.; Yang, S. Multi-scale network toward real-world image denoising. *Int. J. Mach. Learn. Cybern.* **2023**, *14*, 1205–1216. [CrossRef]
3. Xu, S.; Chen, X.; Tang, Y.; Jiang, S.; Cheng, X.; Xiao, N. Learning from multiple instances: A two-stage unsupervised image denoising framework based on deep image prior. *Appl. Sci.* **2022**, *12*, 10767. [CrossRef]
4. Budhiraja, S.; Goyal, B.; Dogra, A.; Agrawal, S. An efficient image denoising scheme for higher noise levels using spatial domain filters. *Biomed. Pharmacol. J.* **2018**, *11*, 625–634.
5. Li, Z.; Liu, H.; Cheng, L.; Jia, X. Image denoising algorithm based on gradient domain guided filtering and NSST. *IEEE Access* **2023**, *11*, 11923–11933. [CrossRef]
6. Aboturab, M.R.; Alfalou, A. Multiple color image fusion, compression, and encryption using compressive sensing, chaotic-biometric keys, and optical fractional Fourier transform. *Opt. Laser Technol.* **2022**, *151*, 108071. [CrossRef]
7. Xu, H.; Jia, X.; Cheng, L.; Huang, H. Affine non-local Bayesian image denoising algorithm. *Vis. Comput.* **2023**, *39*, 99–118. [CrossRef]
8. Dabov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans. Image Process.* **2007**, *16*, 2080–2095. [CrossRef]
9. Chen, F.; Huang, M.; Ma, Z.; Li, Y.; Huang, Q. An iterative weighted-mean filter for removal of high-density salt-and-pepper noise. *Symmetry* **2020**, *12*, 1990. [CrossRef]
10. Lai, R.; Mo, Y.; Liu, Z.; Guan, J. Local and nonlocal steering kernel weighted total variation model for image denoising. *Symmetry* **2019**, *11*, 329. [CrossRef]
11. Li, M.; Cai, G.; Bi, S.; Zhang, X. Improved TV Image Denoising over Inverse Gradient. *Symmetry* **2023**, *15*, 678. [CrossRef]
12. Ou, Y.; Swamy, M.; Luo, J.; Li, B. Single image denoising via multi-scale weighted group sparse coding. *Signal Process.* **2022**, *200*, 108650. [CrossRef]
13. Izadi, S.; Sutton, D.; Hamarneh, G. Image denoising in the deep learning era. *Artif. Intell. Rev.* **2023**, *56*, 5929–5974. [CrossRef]
14. Foerster, J.; Farquhar, G.; Afouras, T.; Nardelli, N.; Whiteson, S. Counterfactual multi-agent policy gradients. *Proc. AAAI Conf. Artif. Intell.* **2018**, *32*. [CrossRef]
15. Varga, B.; Kulcsár, B.; Chehreghani, M.H. Deep Q-learning: A robust control approach. *Int. J. Robust Nonlinear Control.* **2023**, *33*, 526–544. [CrossRef]
16. Kim, Y.; Soh, J.W.; Park, G.Y.; Cho, N.I. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 3482–3492.
17. Zhang, H.; Li, Y.; Chen, H.; Gong, C.; Bai, Z.; Shen, C. Memory-efficient hierarchical neural architecture search for image restoration. *Int. J. Comput. Vis.* **2022**, *130*, 157–178. [CrossRef]
18. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A comprehensive survey on transfer learning. *Proc. IEEE* **2020**, *109*, 43–76. [CrossRef]

19. Valsesia, D.; Fracastoro, G.; Magli, E. Deep graph-convolutional image denoising. *IEEE Trans. Image Process.* **2020**, *29*, 8226–8237. [CrossRef]

20. Kim, D.W.; Ryun Chung, J.; Jung, S.W. Grdn: Grouped residual dense network for real image denoising and gan-based real-world noise modeling. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019.

21. Yu, K.; Wang, X.; Dong, C.; Tang, X.; Loy, C.C. Path-restore: Learning network path selection for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 7078–7092. [CrossRef]

22. Chen, X.; Shen, J. Monte Carlo Noise Reduction Algorithm Based on Deep Neural Network in Efficient Indoor Scene Rendering System. *Adv. Multimed.* **2022**, *2022*, 9169772. [CrossRef]

23. Halidou, A.; Mohamadou, Y.; Ari, A.A.A.; Zacko, E.J.G. Review of wavelet denoising algorithms. *Multimed. Tools Appl.* **2023**, *82*, 41539–41569. [CrossRef]

24. Wang, M.; Wang, S.; Ju, X.; Wang, Y. Image Denoising Method Relying on Iterative Adaptive Weight-Mean Filtering. *Symmetry* **2023**, *15*, 1181. [CrossRef]

25. Teng, L.; Li, H.; Yin, S. Modified pyramid dual tree direction filter-based image denoising via curvature scale and nonlocal mean multigrade remnant filter. *Int. J. Commun. Syst.* **2018**, *31*, e3486. [CrossRef]

26. Wang, Y.; Pang, Z.F. Image denoising based on a new anisotropic mean curvature model. *Inverse Probl. Imaging* **2023**, *17*, 870–889. [CrossRef]

27. Abazari, R.; Lakestani, M. A hybrid denoising algorithm based on shearlet transform method and Yaroslavsky's filter. *Multimed. Tools Appl.* **2018**, *77*, 17829–17851. [CrossRef]

28. Goyal, B.; Dogra, A.; Sangaiah, A.K. An effective nonlocal means image denoising framework based on non-subsampled shearlet transform. *Soft Comput.* **2022**, *26*, 7893–7915. [CrossRef]

29. Liu, C.; Zhang, L. A Novel Denoising Algorithm Based on Wavelet and Non-Local Moment Mean Filtering. *Electronics* **2023**, *12*, 1461. [CrossRef]

30. You, N.; Han, L.; Zhu, D.; Song, W. Research on image denoising in edge detection based on wavelet transform. *Appl. Sci.* **2023**, *13*, 1837. [CrossRef]

31. Al-Shamasneh, A.R.; Ibrahim, R.W. Image Denoising Based on Quantum Calculus of Local Fractional Entropy. *Symmetry* **2023**, *15*, 396. [CrossRef]

32. Kumar, A.; Ahmad, M.O.; Swamy, M. An efficient denoising framework using weighted overlapping group sparsity. *Inf. Sci.* **2018**, *454*, 292–311. [CrossRef]

33. Jia, H.; Yin, Q.; Lu, M. Blind-noise image denoising with block-matching domain transformation filtering and improved guided filtering. *Sci. Rep.* **2022**, *12*, 16195. [CrossRef] [PubMed]

34. Mahdaoui, A.E.; Ouahabi, A.; Moulay, M.S. Image denoising using a compressive sensing approach based on regularization constraints. *Sensors* **2022**, *22*, 2199. [CrossRef]

35. Liu, S.; Hu, Q.; Li, P.; Zhao, J.; Wang, C.; Zhu, Z. Speckle suppression based on sparse representation with non-local priors. *Remote. Sens.* **2018**, *10*, 439. [CrossRef]

36. Bhargava, G.U.; Sivakumar, V.G. An Effective Method for Image Denoising Using Non-local Means and Statistics based Guided Filter in Nonsubsampled Contourlet Domain. *Int. J. Intell. Eng. Syst.* **2019**, *12*. [CrossRef]

37. Qi, G.; Hu, G.; Mazur, N.; Liang, H.; Haner, M. A novel multi-modality image simultaneous denoising and fusion method based on sparse representation. *Computers* **2021**, *10*, 129. [CrossRef]

38. Xie, Z.; Liu, L.; Luo, Z.; Huang, J. Image denoising using nonlocal regularized deep image prior. *Symmetry* **2021**, *13*, 2114. [CrossRef]

39. Fan, L.; Li, H.; Shi, M.; Hua, Z.; Zhang, C. Two-stage image denoising via an enhanced low-rank prior. *J. Sci. Comput.* **2022**, *90*, 57. [CrossRef]

40. Lü, J.; Luo, X.; Qi, S.; Peng, Z. Image denoising using weighted nuclear norm minimization with preserving local structure. *Laser Optoelectron. Prog.* **2019**, *56*, 161006.

41. Buades, A.; Coll, B.; Morel, J.M. Non-local means denoising. *Image Process. Line* **2011**, *1*, 208–212. [CrossRef]

42. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef]

43. Wang, J.; Lu, Y.; Lu, G. Lightweight image denoising network with four-channel interaction transform. *Image Vis. Comput.* **2023**, *137*, 104766. [CrossRef]

44. Yan, H.; Chen, X.; Tan, V.Y.; Yang, W.; Wu, J.; Feng, J. Unsupervised image noise modeling with self-consistent GAN. *arXiv* **2019**, arXiv:1906.05762.

45. Zhao, D.; Ma, L.; Li, S.; Yu, D. End-to-end denoising of dark burst images using recurrent fully convolutional networks. *arXiv* **2019**, arXiv:1904.07483.

46. Yang, J.; Liu, X.; Song, X.; Li, K. Estimation of signal-dependent noise level function using multi-column convolutional neural network. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2418–2422.

47. Yu, S.; Park, B.; Jeong, J. Deep iterative down-up cnn for image denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 15–20 June 2019.

48. Chen, J.; Chen, J.; Chao, H.; Yang, M. Image blind denoising with generative adversarial network based noise modeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3155–3164.

49. Bian, S.; He, X.; Xu, Z.; Zhang, L. Hybrid Dilated Convolution with Attention Mechanisms for Image Denoising. *Electronics* **2023**, *12*, 3770. [CrossRef]

50. Zhang, K.; Zuo, W.; Zhang, L. Learning a single convolutional super-resolution network for multiple degradations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 3262–3271.

51. Shah, V.H.; Dash, P.P. Two stage self-adaptive cognitive neural network for mixed noise removal from medical images. *Multimed. Tools Appl.* **2023**, 1–23. [CrossRef]

52. Obeso, A.M.; Benois-Pineau, J.; Vázquez, M.S.G.; Acosta, A.Á.R. Visual vs internal attention mechanisms in deep neural networks for image classification and object detection. *Pattern Recognit.* **2022**, *123*, 108411. [CrossRef]

53. Anwar, S.; Barnes, N. Real image denoising with feature attention. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Republic of Korea, 27 October–2 November 2019; pp. 3155–3164.

54. Zhang, K.; Zuo, W.; Zhang, L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising. *IEEE Trans. Image Process.* **2018**, *27*, 4608–4622. [CrossRef]

55. Tabassum, S.; Gowre, S.C. Optimal image Denoising using patch-based convolutional neural network architecture. *Multimed. Tools Appl.* **2023**, *82*, 29805–29821. [CrossRef]

56. Mei, Y.; Fan, Y.; Zhou, Y. Image super-resolution with non-local sparse attention. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 3517–3526.

57. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Multi-stage progressive image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 14821–14831.

58. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

59. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.

60. Wang, Y.; Song, X.; Chen, K. Channel and space attention neural network for image denoising. *IEEE Signal Process. Lett.* **2021**, *28*, 424–428 [CrossRef]

61. Zhang, Y.; Li, K.; Li, K.; Sun, G.; Kong, Y.; Fu, Y. Accurate and fast image denoising via attention guided scaling. *IEEE Trans. Image Process.* **2021**, *30*, 6255–6265. [CrossRef]

62. Liu, Y.; Qin, Z.; Anwar, S.; Ji, P.; Kim, D.; Caldwell, S.; Gedeon, T. Invertible denoising network: A light solution for real noise removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 13365–13374.

63. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale residual network for image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 517–532.

64. Chen, L.; Lu, X.; Zhang, J.; Chu, X.; Chen, C. Hinet: Half instance normalization network for image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 182–192.

65. Roy, A.; Singha, J.; Manam, L.; Laskar, R.H. Combination of adaptive vector median filter and weighted mean filter for removal of high-density impulse noise from colour images. *IET Image Process.* **2017**, *11*, 352–361. [CrossRef]

66. Lyu, Q.; Guo, M.; Pei, Z. DeGAN: Mixed noise removal via generative adversarial networks. *Appl. Soft Comput.* **2020**, *95*, 106478. [CrossRef]

67. Malinski, L.; Smolka, B. Self-tuning fast adaptive algorithm for impulsive noise suppression in color images. *J. Real-Time Image Process.* **2020**, *17*, 1067–1087. [CrossRef]

68. Lone, M.R.; Khan, E. A good neighbor is a great blessing: Nearest neighbor filtering method to remove impulse noise. *J. King Saud-Univ.-Comput. Inf. Sci.* **2022**, *34*, 9942–9952. [CrossRef]

69. Lu, X.; Li, F. Fine-tuning convolutional neural network based on relaxed Bayesian-optimized support vector machine for random-valued impulse noise removal. *J. Electron. Imaging* **2023**, *32*, 013006. [CrossRef]

70. Satti, P.; Shrotriya, V.; Garg, B.; Surya Prasath, V. DIBS: Distance-and intensity-based separation filter for high-density impulse noise removal. *Signal Image Video Process.* **2023**, *17*, 4181–4188. [CrossRef]

71. Ri, G.I.; Kim, S.J.; Kim, M.S. Improved BM3D method with modified block-matching and multi-scaled images. *Multimed. Tools Appl.* **2022**, *81*, 12661–12679. [CrossRef]

72. Xu, J.; Zhang, L.; Zhang, D. A trilateral weighted sparse coding scheme for real-world image denoising. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 20–36.

73. Quan, Y.; Chen, M.; Pang, T.; Ji, H. Self2self with dropout: Learning self-supervised denoising from single image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020; pp. 1890–1898.

74. Zoran, D.; Weiss, Y. From learning models of natural image patches to whole image restoration. In Proceedings of the 2011 International Conference on Computer Vision, Barcelona, Spain, 6–13 November 2011; pp. 479–486.

75. Chen, Y.; Pock, T. Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 1256–1272. [CrossRef]

76. Tian, C.; Xu, Y.; Li, Z.; Zuo, W.; Fei, L.; Liu, H. Attention-guided CNN for image denoising. *Neural Netw.* **2020**, *124*, 117–129. [CrossRef]

77. Jia, X.; Liu, S.; Feng, X.; Zhang, L. Focnet: A fractional optimal control network for image denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 15–19 June 2019; pp. 6054–6063.

78. Liang, Z.; Wang, Y.; Wang, L.; Yang, J.; Zhou, S. Light field image super-resolution with transformers. *IEEE Signal Process. Lett.* **2022**, *29*, 563–567. [CrossRef]

79. Zamir, S.W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F.S.; Yang, M.H.; Shao, L. Cycleisp: Real image restoration via improved data synthesis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 14–19 June 2020; pp. 2696–2705.

80. Yue, Z.; Zhao, Q.; Zhang, L.; Meng, D. Dual adversarial network: Toward real-world noise removal and noise generation. In Proceedings of the Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part X 16; Springer: Berlin/Heidelberg, Germany, 2020; pp. 41–58.

81. Kulikov, V.; Yadin, S.; Kleiner, M.; Michaeli, T. Sinddm: A single image denoising diffusion model. In Proceedings of the International Conference on Machine Learning, PMLR, Honolulu, HI, USA, 23–29 July 2023; pp. 17920–17930.

82. Thakur, R.K.; Maji, S.K. Multi scale pixel attention and feature extraction based neural network for image denoising. *Pattern Recognit.* **2023**, *141*, 109603. [CrossRef]

83. Zhang, D.; Zhou, F. Self-supervised image denoising for real-world images with context-aware transformer. *IEEE Access* **2023**, *11*, 14340–14349. [CrossRef]