An Unsupervised Weight Map Generative Network for Pixel-Level Combination of Image Denoisers

Lijia Yu, Jie Luo, Shaoping Xu *, Xiaojun Chen and Nan Xiao

School of Mathematics and Computer Sciences, Nanchang University, Nanchang 330031, China; 5907119011@email.ncu.edu.cn (L.Y.); 406100210094@email.ncu.edu.cn (J.L.); 401030920025@email.ncu.edu.cn (X.C.); 406100210085@email.ncu.edu.cn (N.X.)
* Correspondence: xushaoping@ncu.edu.cn

Abstract: Image denoising is a classic but still important issue in image processing as the denoising effect has a significant impact on subsequent image processing results, such as target recognition and edge detection. In the past few decades, various denoising methods have been proposed, such as model-based and learning-based methods, and they have achieved promising results. However, no stand-alone method consistently outperforms the others in different complex imaging situations. Based on the complementary strengths of model-based and learning-based methods, in this study, we design a pixel-level image combination strategy to leverage their respective advantages for the denoised images (referred to as initial denoised images) generated by individual denoisers. The key to this combination strategy is to generate a corresponding weight map of the same size for each initial denoised image. To this end, we introduce an unsupervised weight map generative network that adjusts its parameters to generate a weight map for each initial denoised image under the guidance of our designed loss function. Using the weight maps, we are able to fully utilize the internal and external information of various denoising methods at a finer granularity, ensuring that the final combined image is close to the optimal. To the best of our knowledge, our enhancement method of combining denoised images at the pixel level is the first proposed in the image combination field. Extensive experiments demonstrate that the proposed method shows superior performance, both quantitatively and visually, and stronger generalization. Specifically, in comparison with the stand-alone denoising methods FFDNet and BM3D, our method improves the average peak signal-to-noise ratio (PSNR) by 0.18 dB to 0.83 dB on two benchmarking datasets crossing different noise levels. Its denoising effect is also greater than other competitive stand-alone methods and combination methods, and has surpassed the denoising effect of the second-best method by 0.03 dB to 1.42 dB. It should be noted that since our image combination strategy is generic, the proposed combined strategy can not only be used for image denoising but can also be extended to low-light image enhancement, image deblurring or image super-resolution.

Keywords: image denoising; noise reduction effect; image combination; unsupervised learning; pixel-level combination

1. Introduction

With the continuous spread of large numbers of images on the Internet due to the rapid development of mobile Internet and social networks, there is an increasing demand for more accurate, high-quality images. However, during the acquisition, compression and transmission of images, the image quality is inevitably deteriorated due to noise [1–3]. Therefore, image denoising is applied in modern image processing systems [4]. This method aims to remove noise from a noisy image while minimizing the differences between the original image and the denoised image [5]. Various researchers have continuously investigated this technique but it remains challenging as it is a well-known ill-posed...
problem with no unique solution from a mathematical standpoint [6–8]. Generally, image denoising can be modeled mathematically as follows:

\[ y = x + n, \]  

where \( y \) is the noisy image, \( x \) is the clean image, and \( n \) is the noise, which is usually modeled using standard deviation \( \sigma \) as additive white Gaussian noise (AWGN) [9]. In the past few decades, a variety of denoising methods have been proposed, which can be categorized into two groups: model-based and learning-based denoising methods [10].

In earlier years, the typical methods were model-based. Earlier model-based methods rely on the statistical analysis of the internal information of the given images. Representative examples of model-based methods are non-local means (NLM) [11] and block matching and 3D filtering (BM3D) [12]. In 2005, Baudes et al. first utilized non-local self-similarity in images and proposed NLM. Based on the observation that a natural image patch includes several similar counterparts inside the same image, NLM removes the noise by stacking non-local self-similar patches, which works well with images that have repeating patterns in spatially diverse regions. However, it is computationally impossible to search for similar patches within the whole image; hence, usually, a small vicinity of the patch is explored while looking for optimal matches [13]. Later, inspired by the idea of NLM, Dabov et al. proposed BM3D, which included a two-step denoising process. First, the input image was denoised and a basic estimate of image blocks was generated. Then, by using collaborative filtering of the generated estimate, the noise reduction effect was increased. However, the block-wise process may result in distorted output. For instance, when the noise level rises gradually, the denoising performance of the model decreases, and artifacts emerge, particularly in flat regions. In 2011, Dong et al. presented a non-locally centralized sparse representation (NCSR) [14]. It makes use of a sparse representation of image patches and non-local similarity, altering the goal of image denoising to suppressing sparse coding noise. The NCSR has demonstrated remarkable performance for natural images with both smooth and textured regions. However, the algorithm is not suitable for many applications owing to its computational complexity in non-local estimates of unknown sparse coefficients. Although these model-based methods have fundamental mathematical derivations, they still have a few drawbacks. First, they are frequently time-consuming owing to the complexity of the iterative optimization processes. Second, only the internal information of the noisy image is used by the model-based methods. No external information is used; this leads to the denoising effect being considerably impaired while recovering flat regions and texture patterns with excessive noise [15].

With the rapid development of deep learning, learning-based methods have shown great potential in image denoising [16]. The learning-based methods aim to train a deep denoising network with datasets and learn the noise model implicit mapping from the noisy to the clean images to form external priors [2]. In recent years, the popularity of learning-based methods has surpassed that of model-based methods. In 2016, Zhang et al. suggested a feed-forward denoising convolutional neural network (DnCNN) [17]. They incorporated residual learning [18] and batch normalization [19] to speed up the training process while improving denoising outcomes. With the supervision of noisy–clean paired images, it achieved superior performance compared to traditional methods. However, the model can only handle noisy images well when the noise level is in the preset range; hence, it lacks generalization capability. To alleviate the problem in DnCNN of a lack of flexibility, Zhang et al. introduced a flexible and fast CNN solution for CNN-based image denoising called FFDNet [20]. It contains an adjustable noise level map and can be trained to recognize the noise level. Experiments revealed that the method can handle several images with different noise levels, with high performance and efficiency. However, the denoising performance of FFDNet is limited to the training dataset, indicating that it can only obtain excellent results on noisy images that are similar to the images of the training set. In other words, FFDNet may generalize poor results for input images that are corrupted with complex noise, especially real-world noisy images. To improve the generalization capability of the deep
CNN denoiser, in 2019, Guo et al. used real-synthetic noise and real-image noise to train a convolutional blind denoising network (CBDNet) [21]. Although CBDNet performed well under different noise levels, it is less efficient and flexible owing to its separate subnets and the need for manual intervention. Soon after, Yue et al. presented a variational denoising network (VDNet) [22]. It automatically estimates the noise distribution and the noise-free image inside a unique Bayesian framework using variational inference methods. The VDNet surpassed CBDNet in both performance and generalization capability. However, it does not consider signal autocorrelation, making it difficult to estimate the real-world noise in some forms. In 2022, for the enhanced control and interpretability of a learning-based denoiser, Liang et al. proposed a novel architecture based on a denoising network called controllable confidence-based image denoising (CCID) [23]. CCID merges the deep denoising network with a reliable filter to reduce artifacts. The results revealed that CCID improves the interpretability and control and outperforms the former deep denoisers, such as DnCNN, especially when the test data differ from the training data. Compared to the model-based denoising methods, learning-based methods are advantageous in the following two aspects. (1) They benefit from numerous datasets. These learning-based methods make full use of external information and demonstrate an extremely competitive denoising effect. (2) Although the learning-based methods need to utilize their network to extract features, which is time-consuming, the inference of CNN is very efficient at the test stage owing to the computation ability of the GPU. Extensive experiments demonstrate that their processing speed is mostly higher than that of model-based methods [24]. However, their strong dependence on data results in certain drawbacks. Owing to their strong dependence on external information, the performance of learning-based methods is limited to their training data [25]. These methods find it difficult to deal with universal image denoising problems because they are not adaptive to a given noisy image with different noise levels, image texture and scene classification from the dataset; some fine-scale image details may not be well preserved. Therefore, for the learning-based models, their strong data dependence tends to result in poor generalization capability and flexibility [26].

Although various image denoising methods have been proposed and have achieved highly promising results over the last decade, none of the denoising methods is perfect owing to their technical characteristics. We can safely state that it is difficult for an individual denoiser to promote the denoising performance better than others under various conditions. Conversely, model-based methods and learning-based methods are complementary to each other. For example, model-based methods are adaptable to strong self-similarity images [27], whereas the learning-based methods need cumbersome training to learn a model before the test stage and tend to be restricted by certain images. As a result, various strategies have been proposed to investigate a combination of these methods for denoising, leveraging their respective advantages. To investigate their integration, which uses their respective merits, Choi et al. [28] considered a combination of the denoised images obtained using these methods and then proposed a Consensus Neural Network (CsNet). However, there is still scope for further improvement. For example, it ignores the fact that each initial denoiser contains denoising characteristics regarding the details of local areas. Therefore, it is unsuitable to estimate the weight for each initial denoised image. Therefore, this study aims to further enhance the denoising effect. We propose a weight map generative network for the pixel-level image combination of initial denoised images to obtain an optimal, well-combined image. The contributions of this study are as follows:

1. Compared to CsNet [28], which sets a stand-alone weight for each initial denoised image, we introduce a finer combining granularity scale in this study. Specifically, we use a deep learning network to set appropriate weights for each pixel, which contain information about the contribution of the initial denoised images on the combined image, ensuring that it can retain the details well. We call this a pixel-level combination strategy.

2. Conventional supervised CNN-based denoising methods need noisy–clean paired images to achieve high performance. In our study, we use an unsupervised learning
method to achieve the optimal weight maps without noisy–clean paired images [29]. It reduces the cost of collecting considerable training pairs for training and makes the generalization capability of our method excellent.

3. The proposed method can be easily extended, allowing the free combination of several initial denoised images generated by any denoiser. Therefore, in our study, we performed combination experiments with different classes as well as different numbers of denoised images to find the complementarity between different denoising methods. Currently, researchers are attempting to improve the denoising performance. As more efficient denoising methods are proposed, in the future, our approach will still perform well. We can employ several different methods that are more efficient and complementary to each other to enhance the denoising effect.

The remainder of this study is structured as follows: Section 2 presents image-level and patch-level combination strategy models relevant to our study. In Section 3, we outline the image combination problem and introduce the proposed pixel-level combination strategy. In Section 4, we present our datasets, experimental setup and experimental results, including the selection of combinatorial patterns, ablation experiments and comparison with other state-of-the-art denoising methods. Finally, this study is concluded in Section 5.

2. Related Work

Generally, the granularity of image combinations can be divided into three levels: image level, patch level and pixel level. Theoretically, the smaller the granularity of the combination, the higher the potential for denoising effect enhancement. Our method belongs to the pixel level and has the potential to obtain optimal combination results. To help the reader gain a better understanding of the background, in this section, we briefly review both the image-level image combination strategy and the patch-level image combination strategy. We first introduce the image-level combination strategy, CsNet. Subsequently, a patch-level combination strategy called structural patch decomposition (SPD) is reviewed.

2.1. Consensus Neural Network

In 2019, CsNet was proposed by Choi et al. They observed that denoised images using different denoising methods are significantly complementary, and there may be a way to combine the denoised images to generate an excellent overall result. Therefore, they integrated different denoisers in the CsNet to generate a combined denoised image, which is better than those obtained by the individual denoising techniques. It mainly handles Gaussian noise. As can be seen in Figure 1, it can be divided into two stages: the combining stage and the boosting stage.

The first stage is employed for preliminary estimates to obtain the combined image. For simplicity, we used the same notation as in [28]. A set of initial estimates \( \{ \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_K \} \) can be derived using \( K \) image denoisers \( D_1, D_2, \ldots, D_K \). To concatenate these initial estimates, the matrix \( \hat{X} = [\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_K] \in \mathbb{R}^{N \times K} \) is constructed. For image combination, CsNet focuses on the linear combination of estimators, which is used to keep the model unbiased and ensure lower variance. That is, given \( \hat{X} \), the linearly combined estimate can be obtained as

\[
\hat{x} = \sum_{k=1}^{K} w_k \hat{x}_k = \hat{X}w
\]

where \( w \) is the vector of combination weights. To identify the weights \( w_1, w_2, \ldots, w_K \), Choi et al. used a convex optimal combination framework based on the previous work of Jaggi [30] to solve a convex optimization problem.
Figure 1. The architecture of CsNet [28]. $D_1, D_2, \ldots, D_K$ are the initial denoisers, $M$ refers to the estimator used to estimate the mean squared error (MSE) of each initial denoised image, $w_1, w_2, \ldots, w_K$ are the generated weights, which are determined by the solver by finding a solution to a convex optimization problem.

Using the estimator $M$, the CsNet method estimates a set of typical denoisers to generate their respective mean squared error (MSE) [31] of the initial denoised image. The MSE between the ground truth $x \in \mathbb{R}^N$ and the combined estimate $\hat{x} \in \mathbb{R}^N$ is

$$\text{MSE}(\hat{x}, x) \overset{\text{def}}{=} \mathbb{E}[||\hat{x} - x||^2] = \mathbb{E}[||\hat{X}w - x||^2] \quad (3)$$

According to the MSE of each denoised image, the optimal combination problem can be solved by obtaining the weight vector $w \in \mathbb{R}^K$ and minimizing the MSE. Subsequently, depending on the optimal weight, it composites the denoised images together.

$$\min_w \mathbb{E}[||\hat{X}w - x||^2] \quad \text{subject to} \quad w^T1 = 1, \quad \text{and} \quad w \succeq 0 \quad (4)$$

where the constraints $w^T1 = 1$ and $w \succeq 0$ ensure that the total weights are 1 and the combined estimate is nonnegative, respectively.

However, after the first stage, the image quality of the combined image is not excellent. To enhance the contrast and recover the lost features of the combined image, a deep learning-based booster was used at the end of the CsNet method, where variables were updated one by one in the iterations. Specifically, based on the current estimate $\hat{x}^{(t)}$ and the observed image $y$, the next estimate $\hat{x}^{(t+1)}$ can be obtained by the equation

$$\hat{x}^{(t+1)} = B_t\left(y, \hat{x}^{(t)}\right) + \hat{x}^{(t)} \quad (5)$$

Here, $B_t$ is a multi-layer convolutional neural network that implements nonlinear mapping. In the CsNet model, $B_t$ is composed of three convolutional layers and three deconvolutional layers, and the size of the convolution kernels is $3 \times 3$. The input of the network is the pair $(y, \hat{x}^{(t)})$. In addition, Choi et al. provided skip connections to ensure that information is not lost while passing through the network layer. To obtain a better improvement effect, the CsNet model uses five cascaded networks $B_t(t = 1, 2, \ldots, T)$, where $T$ is 3.

CsNet uses multiple denoisers to improve the output image, which has superior denoising performance compared to both model-based and learning-based denoisers. It contributes to the solution of noise level mismatch, and the selection and combination of different denoiser types, but still displays coarse granularity, which has an impact on the quality of the final combined images. Thus, it must employ the booster module to optimize and improve the preliminary optimized image $\hat{z}$ again, with the enhancement
process repeated numerous times to obtain the final high-quality noise reduction effect. Moreover, the denoising effect of CsNet is mainly dominated by the MSE estimator. However, the MSE estimator does not perform well in terms of accuracy and it suppresses any further improvement in the image quality. In general, the implementation process of CsNet is complicated, resulting in a long execution time and limited improvement in the image quality.

2.2. Structural Patch Decomposition Approach

In addition to the image-level combination strategy, image combination can also be performed at the patch level. The patch-level combination strategy improves the granularity of image processing and retains more available details in images compared to the image-level combination strategy. In recent years, various researchers have proposed their patch-level fusion schemes, among which Ma et al. proposed SPD for multi-exposure image fusion in [32]. Although the SPD was not originally used in image denoising, we found that applying it to denoised image combination yielded higher-quality results compared to typical methods, such as CsNet. Therefore, in this section, we present a discussion on structural patch decomposition.

Initially, the SPD approach uses a moving window with a fixed stride to extract $K$ overlapping image patch sequences $\{p_k\} = \{p_k | k = 1, \ldots, K\}$ from the entire image. They decompose a given patch $p_k$ into three parts, such as signal strength, signal structure and mean intensity. Its decomposition formula is

$$p_k = \|p_k - \mu_{p_k}\| \cdot \frac{p_k - \mu_{p_k}}{\|p_k - \mu_{p_k}\|} + \mu_{p_k}$$

$$= \|\tilde{p}_k\| \cdot \frac{\tilde{p}_k}{\|\tilde{p}_k\|} + \mu_{p_k}$$

$$= c_k \cdot s_k + l_k$$

where $\|\cdot\|$ is the $l_2$ norm of a vector, $\mu_{p_k}$ is the mean value of the patch, and $\tilde{p}_k = p_k - \mu_{p_k}$ is a mean-removed patch. As a result, the scalar $c_k = \|\tilde{p}_k\|$, $l_k = \mu_{p_k}$ and vector $s = \frac{\tilde{p}_k}{\|\tilde{p}_k\|}$ have physical significances that correspond to the strength component, mean intensity component and structure component of $p_k$, respectively.

First, the strength component is analyzed. Here, the contrast of the combined patch is determined by the highest contrast. It is expressed as

$$\hat{c} = \max_{1 \leq k \leq K} c_k = \max_{1 \leq k \leq K} \|\tilde{p}_k\|$$

Regarding the structure component, to make the combined image patch structure reflect all source image patch structures, a basic implementation of this relationship is provided

$$\hat{s} = \frac{s}{\|s\|} \quad \text{and} \quad s = \frac{\sum_{k=1}^{K} S(\tilde{p}_k) s_k}{\sum_{k=1}^{K} S(\tilde{p}_k)}$$

where $S(\cdot)$ is a weighting function given by $S(\tilde{p}_k) = \|\tilde{p}_k\|^p$ that defines the contribution factor for each source image patch.

For the mean intensity component, it is denoted as

$$\hat{l} = \frac{\sum_{k=1}^{K} L(\mu_{p_k}, l_k) l_k}{\sum_{k=1}^{K} L(\mu_{p_k}, l_k)}$$

where $L(\cdot)$ is a weighting function. It contains two parameters, including the local mean value of the current patch and the global mean value $\mu_{p_k}$ of the image, as inputs. $L(\cdot)$ measures the exposures of $p_k$ in $x_k$, imposing a high penalty when the image and/or the
current patch are under/over-exposed. After determining the \( \hat{c}, \hat{s}, \text{and} \hat{l} \) components using Equations (7)–(9), respectively, the reconstructed patch can be expressed as

\[
\hat{x} = \hat{c} \cdot \hat{s} + \hat{l} \quad (10)
\]

After the overlapping image patches are reconstructed, they are placed back into the entire image and the pixels in the overlapping patches are averaged to produce a great overall result.

In our study, we apply SPD to image denoising. For the co-located initial denoised patches produced by stand-alone methods, SPD decomposes them into signal intensity, signal structure and average intensity. Then, it combines the initial denoised patches by evaluating these three components. The combined patch always retains rich details, image structure information and intensity information. Finally, these combined patches are placed back into the entire denoised image. As a result, in comparison with image-level combination methods, SPD uses the image local information more effectively during combination and produces a higher-quality denoised image.

However, some artifacts around edges are produced because of the inability of the model to create a sufficiently smooth transition between exposures near strong edges. Therefore, the patch-level combination strategy is still not optimal and there is scope for improvement. In addition, the weights of each component of the SPD method are handcrafted, which does not ensure that the optimal combination results will always be achieved in all parts of the image. Hence, the patch-level combination strategy is still rough and has room for improvement.

3. Methodology

In this section, we briefly look at the image combination problem and explain the processing and structure of our proposed pixel-level combination strategy.

3.1. Basic Concept

We illustrate below, with an example, the complementary nature of the model- and learning-based image denoising methods. Figure 2 shows the model- and learning-based denoisers called BM3D and FFDNet, respectively. These images are corrupted by a medium level of noise (\( \sigma = 30 \)). For better visual comparison, a particular region is enlarged for images. Based on the magnified region in the Barbara images, the BM3D excels in areas with repeating textures. In contrast, in the Boat image, we can observe that FFDNet is better in generic content than BM3D because FFDNet has learned image priors from external data, which is specific to the image to be denoised. These results show that BM3D is successful at retaining the repeating textures, whereas FFDNet is capable of removing noise in the generic content. Therefore, FFDNet and BM3D are complementary to each other. To further leverage each of their advantages, we aimed to combine the two results.

Moreover, we introduced a pixel-level combination strategy, which makes full use of the complementarity among different denoisers, to retain the well-denoised details of each preprocessed denoised image in finer granularity. Therefore, the problem of improving the denoising effect in this study is a pixel-level image combination problem, which typically follows a weighted summation framework that combines the \((i, j)\)-th pixel \( \hat{x}_k(i, j) \) in image \( x_k \) with the corresponding weighted \( w_k(i, j) \) in weight map \( W_k \) to achieve an optimal combined result \( \hat{x}(i, j) \). Specifically, in our study, the aforementioned input images are the initial denoised images \( \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k \) generated by different denoisers \( D_1, D_2, \ldots, D_K \). Thus, for each pixel, our study can be described mathematically as

\[
\hat{x}(i, j) = \sum_{k=1}^{K} w_k(i, j) \hat{x}_k(i, j) \quad (11)
\]

Considering that these initial denoised images are complementary to each other, we set weights for all pixels of each denoised image, where the sum of the weights in the same
position is 1 after normalization. The weight map of each denoised image is difficult to be set through manual designing since it is complex and different. Hence, we used the unsupervised deep learning method to find these optimal weight maps. In other words, we intended to use the unsupervised deep learning network to generate a normalized weight map for each denoised image. The normalization operation ensures that the intensity value of the combined images remains within the 0 to 1 or 0 to 255 ranges. The normalization formula is defined as:

$$w_k(i, j) = \frac{o_k(i, j)}{\sum_{n=1}^{N} o_n(i, j)}$$

where $o_k(i, j)$ represents the original weight of the $(i,j)$-th pixel of the $k$-th initial denoised image generated by our unsupervised deep learning network, and $w_k(i, j)$ represents the $(i,j)$-th pixel of the $k$-th initial denoised image after normalization.

In summary, we used an unsupervised deep learning network to generate a weight map for each initial denoised image under the guidance of our designed loss function, where we combined these images at the pixel level to provide the optimal denoising result.

![Figure 2. Denoising results of Barbara and Boat with noise level $\sigma = 30$.](image)

3.2. Unsupervised Weight Map Generative Network

In the Section 2, we have mentioned the image-level combination strategy CsNet. CsNet suffers from coarse combination granularity and many details of the images cannot be retained well. To compensate for the lost details of the combined image, Choi et al. designed a boosting procedure that complicates its network structure. For the patch-level combination strategy SPD, it is superior to CsNet in terms of detail retention; however, it produces some artifacts around the edges. To solve the above problem in image combination, we aim to develop a new image combination strategy. Unlike existing coarse-grained image combination strategies that focus on images or patches [28,32], we focus on a finer-grained pixel-level image combination strategy. We argue that with a powerful, adaptive unsupervised learning network, we can set weights for each pixel of the image to be combined. To achieve this, in this section, we will present a model for pixel-level image combination, called the unsupervised weight map generative network (UWMGN). The
pipeline of the proposed UWMGN is shown in Figure 3. It generates a corresponding weight map of the same size for each initial denoised image; using the weight maps, we will combine the initial images and then continuously adjust the network parameters under the guidance of the loss function so that the output weight maps can combine to yield an optimal result. Compared to CsNet, our method only requires one stage to complete the denoising process, without an extra boost.

Figure 3. Pipeline of the proposed UWMGN. $D_1, D_2, \ldots, D_K$ are the image denoisers, $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_K$ refer to the initial denoised images, $W_1, W_2, \ldots, W_K$ are the weight maps.

First, given a noisy image $y$, we obtained a set of initial denoised images processed with different denoising methods. Then, the initial denoised images were matched with the corresponding weight maps of the same size, one by one, and each initial denoised image was weighed and combined to obtain the final combined output. To generate weight maps, we designed a backbone network. It uses random noise as the input to generate weight maps without the need for prior training pairs. As the initial input of the backbone network is random noise, its output weight map also begins with random values. We used a novel loss function that iteratively updates the parameters in the generative network and adjusts the convergence point of our model to ensure that the output combined image is close to optimal. Therefore, for every denoised image $\{\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_K\} \in \mathbb{R}^N$, we constructed the corresponding weight maps $\{W_1, W_2, \ldots, W_K\} \in \mathbb{R}^N$. We will discuss the structure of the backbone network and loss function in detail in the following subsections.

3.3. Backbone Network

Recent developments in CNNs for image denoising demonstrate that unsupervised learning networks can achieve good denoising results without requiring datasets with a considerable amount of target images for training [33]. For this purpose, we designed a backbone network. As shown in Figure 4, the backbone network is based on an unsupervised network with an encoder–decoder structure, which is similar to that of the typical U-Net [34], except for the input and output layers. As a result, the main operators in our network are a combination of convolution, batch normalization and ReLu. The left and right sides of our network are the encoder and decoder, respectively. The encoder utilizes a successive convolution layer to extract precise information, and the decoder uses a successive deconvolution layer to recover the information. The advantage of the typical U-Net encoder–decoder network structure is that it uses a multi-scale processing method...
to expand the processing field of view, which greatly reduces the memory requirements for storing parameters. Simultaneously, the connection between the same level features in the encoder–decoder is established through skip connections to minimize information loss due to upsampling and downsampling. It is represented by a dashed arrow in the figure. Compared to the standard U-Net network, for the input of the network structure, our input is not an image but random noise of the same size as the initial denoised images. Since our desired output is $K$ weight maps corresponding to the denoised images, we redesigned the loss function of the U-Net network. In this way, the output weight maps can retain the best part of the corresponding denoised images. This function was derived from the preliminary corrupted image and the combined image, ensuring that our output evolves in a reasonable direction—that is, the noise of the image continuously decreases while the image remains close to the input image without distortion. In the following section, we will analyze the loss function. Lastly, due to the great extensibility of our network, it can generate weight maps for colored images after expanding it by the number of channels.

![Figure 4. Structure of our backbone network. Z refers to the initial random noise. $O_1, O_2, \ldots, O_K$ are the generated original weight maps. $W_1, W_2, \ldots, W_K$ are the weight maps after normalization.](image)

### 3.4. Loss Function

The UWMGN utilizes the loss function as the training target to assess the quality of the combined image without the ground truth. We discuss our loss function in this section.

Information entropy (IE) is a method for evaluating the information content of images. The higher the information entropy, the more details there are in the image and the better the quality of the image. As a result, information entropy loss is used to suppress noise interference. The expression of the information entropy loss function is as follows:

$$
L_{IE} = - \sum_{g=0}^{d} p(g) \log_2 p(g)
$$

(13)

where $p(g)$ denotes the probability of the occurrence of the intensity value $g$ in the combined image $\hat{x}$, and $d$ represents the range of intensity values in the combined image, which usually ranges from 0 to 1 or 0 to 255. We set the information entropy to a negative number such that the $L_{IE}$ amount of the combined image will continue decreasing and more details will be retained, further improving the quality of the final combined image.

Total variation (TV) calculates a value based on the input signal. It was observed by Rudin et al. that the TV of a noisy image is much higher than that of a noise-free image. Therefore, image denoising can be performed by minimizing the TV value [35]. Hence, we utilized TV to regularize the loss function, which will be minimized by CNN to reduce noise in the image. The loss function is given as

$$
L_{TV} = \sum_{i=1}^{H} \sum_{j=1}^{W} \left[ (\hat{x}(i,j+1) - \hat{x}(i,j))^2 + (\hat{x}(i+1,j) - \hat{x}(i,j))^2 \right]^{1/2},
$$

(14)

where $W$ and $H$ are the width and height of the combined image, respectively, and $\hat{x}(i,j)$ is the $(i,j)$-th pixel in the combined image.
As the PSNR [36] is widely used to assess denoising performance, we added an MSE to the loss function to adjust our mapping parameters. MSE is a measure that reflects the difference between the estimated and actual values. In this study, we used MSE loss to measure the difference between the initial noisy image \( y \) and the output image \( \hat{x} \). Then, the loss function is given as \( L_{\text{MSE}} \) in Equation (15).

\[
L_{\text{MSE}} = \frac{1}{WH} \sum_{i=1}^{H} \sum_{j=1}^{W} |y(i, j) - \hat{x}(i, j)|,
\]

where \( y(i, j) \) is the \((i, j)\)-th pixel in the noisy image. In comparison to the standard MSE, Equation (15) can significantly improve the smoothness and contrast of the uniform area of the image.

We obtained the total loss function by combining all the aforementioned loss functions. The total loss is indicated by \( L \).

\[
L = \lambda_{\text{MSE}} L_{\text{MSE}} + \lambda_{\text{IE}} L_{\text{IE}} + \lambda_{\text{TV}} L_{\text{TV}},
\]

where the parameters of \( \lambda_{\text{MSE}} \), \( \lambda_{\text{IE}} \) and \( \lambda_{\text{TV}} \) are set to 1000, 1 and 10, respectively, in the experiments. We emphasize that, based on the design of the aforementioned loss function, the proposed weight convolutional neural network can be trained in an unsupervised manner. It combines the advantages of \( L_{\text{IE}} \), \( L_{\text{TV}} \) and \( L_{\text{MSE}} \). Here, \( L_{\text{IE}} \) and \( L_{\text{TV}} \) ensure that the obtained output weight maps can reduce noise interference to obtain the optimal quality of combined image, and \( L_{\text{MSE}} \) effectively reduces image artifacts to guarantee that the fidelity term makes the output image close to the ground truth.

4. Experiments

In this section, we present our datasets, the experimental setup and the experimental results. Specifically, we first perform the selection of the combined patterns and the ablation experiments, and then compare our method to other typical stand-alone denoising methods as well as image combination denoising methods. To evaluate the extensibility of our method, we also perform a low-light image enhancement experiment. The code is available at https://github.com/NCU-YLJ/UWMGN (accessed on 13 September 2021).

4.1. Datasets and Experimental Setup

To evaluate the overall denoising performance of the proposed denoising model comprehensively, we compared our proposed model to ten other models. Some of these competitive methods are used as benchmark methods in the field of image denoising and show good denoising results, such as BM3D [12], weighted nuclear norm minimization (WNNM) [27], NCSR [14], FFDNet [20], DnCNN [17] and VDNet [22]. Specifically, BM3D is a non-locally collaborative filtering method, which performs well in repeated textures. On the other hand, WNNM is a low-rank-based method, and NCSR is a sparse coding scheme. They both give superior results in the homogeneous regions. FFDNet and DnCNN are both representatives of learning-based denoising methods. The learning-based denoising methods aim to learn a mapping between noisy images and denoised images with the assistance of external information, and they usually perform well in noisy images with similar features to the training data. We also selected VDNet because it demonstrates higher generalization capability and flexibility than the first two learning-based models for its explicit form of posterior expression. There is also the combined denoising method CSNet and some state-of-the-art image combination methods that can be used for image denoising, such as SPD [32], unified unsupervised image fusion network (U2Fusion) [37] and a fast multi-exposure image fusion method, namely MEF-Net [38]. We conducted denoising experiments on two benchmark datasets. The two benchmark datasets are as follows: (1) the common image set shown in Figure 5, which consists of 10 test images widely used in various studies, including images with a size of 256 \( \times \) 256 (Monarch, House,
Cameraman and Peppers) and six images with a size of 512 × 512 (Lena, Barbara, Boat, Hill, Couple and Man); (2) 50 randomly selected images from the Berkeley segmentation dataset (BSD) [39], for which 10 representative images are shown in Figure 6. The images in the BSD contain rich content and texture details, which are very suitable for testing the robustness against images of our denoising model. In addition, we visually evaluated the results to subjectively assess the effects of denoising. All models were run on the same hardware platform (i.e., Intel\(\text{R}\) Xeon\(\text{R}\) CPU E5-1603 v4 @ 2.80GHz RAM 16GB, NVIDIA GPU Quadro M4000) and software environment (i.e., Window10 operating system, Pytorch 1.1.0).

Figure 5. A range of 10 commonly used test images in the references.

Figure 6. Some representative images in the BSD.

4.2. Selection of Combination Patterns

Theoretically, our strategy is extensible in terms of the number and type of preprocessing denoisers. Therefore, our first problem is to establish the type and number of denoised methods involved in our combination, which will directly affect the performance of our method. We first conducted combination experiments for two types of denoised images to achieve a trade-off between effect and efficiency. As there are various methods in the field of image denoising, we only selected six representative and high-performing denoisers of various denoising methods to combine. We used the PSNR as our image quality measure to objectively evaluate the performance of the denoising methods. PSNR measures the intensity similarity between the clean image and the denoised image. The higher the PSNR value, the more similar are the intensities of the clean image and the denoised image. Table 1 lists all PSNRs for each method on the common image set. The denoising effect of the combination of FFDNet and BM3D is better than that of the combination of other state-of-the-art denoisers. As we know, BM3D is a typical model-based denoiser, whereas FFDNet is a typical learning-based denoiser. Their combination can explore both the internal and external information of the initial denoised images because of their strong complementarity.
After this, we utilized many combinations of denoising methods, and the results are shown in Table 2. It can be observed that when the number of denoisers is increased by three or more, the improvement in the processing effect of our method is not very significant, and there is even a slight decrease. Therefore, since the preprocessing of the generation of the initial denoised images will take some time and the improvement in the quality of the combined image is also lower, we selected BM3D and FFDNet as our initial denoisers because there is a good trade-off between effect and efficiency. In the experiments, the average processing time for this combination pattern is 4.3 min for each image, which is acceptable. In the future, we can select other state-of-the-art complementary denoisers to generate initial denoised images with specific demands on the computational complexity, denoising effect or a trade-off between the two goals. In our study, we aim to further boost the denoising performance.

### Table 1. Average performance of various combinations on 10 commonly used images with noise level 30.

| Denoisers | FFDNet | DnCNN | NCSR | VDNet | WNNM |
|-----------|--------|-------|------|-------|------|
| BM3D      | 30.82  | 30.60 | 30.36| 30.31 | 30.52|
| FFDNet    | ——     | 29.82 | 30.54| 30.48 | 30.00|
| DnCNN     | ——     | ——   | 30.50| 30.45 | 29.93|
| NCSR      | ——     | ——   | ——  | 30.23 | 30.37|
| VDNet     | ——     | ——   | ——  | ——   | 30.44|

### Table 2. Average performance of different numbers of combinations on 10 commonly used images with noise level 30.

| Number of Denoisers | 2   | 3   | 4   | 5   | 6   |
|---------------------|-----|-----|-----|-----|-----|
| BM3D                |✓    |✓    |✓    |✓    |✓    |
| FFDNet              |✓    |✓    |✓    |✓    |✓    |
| DnCNN               |✓    |✓    |✓    |✓    |✓    |
| WNNM                |✓    |✓    |✓    |✓    |✓    |
| NCSR                |✓    |✓    |✓    |✓    |✓    |
| VDNet               |✓    |✓    |✓    |✓    |✓    |

Table 3. Average PSNR values of different combinations of denoising methods on 10 commonly used images with noise level 30.

### 4.3. Ablation Experiments

To demonstrate the effectiveness of each component proposed in our loss function, we conducted several ablation experiments. The experiments were performed on 10 commonly used test images corrupted by different noise levels, and the results are shown in Table 3 in terms of PSNR. Results that have the highest PSNR values are in bold. Specifically, we designed three experiments by removing two components from the loss function. It can be observed that the MSE loss function contributes most to the loss function, and this is demonstrated by the fact that the PSNR values of the MSE loss function are the highest when compared with the PSNR produced by the loss functions of MSE, IE and TV. We observed that combining these three loss function components together can produce better results. In summary, according to the experimental results, the MSE, IE and TV loss functions are required to obtain a better combination of results.
Table 3. Average PSNR values obtained by selecting different combinations of loss functions on 10 commonly used test images corrupted by different noise levels.

| Noise Level | $L_{IE}$ | $L_{TV}$ | $L_{MSE}$ | $L_{MSE} + L_{TV} + L_{IE}$ |
|-------------|----------|----------|-----------|---------------------------------|
| $\sigma = 10$ | 35.50    | 35.17    | 36.44     | 36.46                           |
| $\sigma = 30$ | 30.03    | 29.85    | 30.81     | 30.82                           |
| $\sigma = 50$ | 27.79    | 27.48    | 28.45     | 28.49                           |
| Average     | 31.11    | 30.83    | 31.90     | 31.92                           |

4.4. Comparison with Other Combination Strategies

In this subsection, we list the experimental results of the proposed pixel-level combination strategy and other combination strategies on 10 commonly used test images and 50 images randomly chosen from the BSD, and highlight the highest PSNR values for each noise level in bold. In addition to CsNet, which we mentioned in our related works, we added three state-of-the-art image combination methods that can be applied to image denoising, to validate the effectiveness and generalization ability of our method. The first is SPD, which we reviewed in the Related Works section. The second is U2Fusion, proposed by Xu et al. Similar to our method, U2Fusion is also an unsupervised end-to-end image combination method, which has great effectiveness and extensibility. U2Fusion performs well in solving multi-focus, multi-exposure and multi-modal image combination problems. Another is MEFNet, proposed by Ma et al., which generates low-resolution weight maps of the initial image using deep learning methods and then combines the initial image with the weight maps to obtain the final result. The average PSNR results with different noise levels ranging from 10 to 60 with a step of 10 are listed in Table 4. Based on the table, our method outperformed the other four methods and achieved the highest average PSNR scores for all cases. On the 10 commonly used test images, our method outperformed even the second-best method by 0.23 to 1.42 dB, which is a significant improvement. On the more complex BSD dataset, our method is still superior to that of the other four methods, and it increased the PSNR result by 0.03 to 5.75 dB, confirming that the proposed pixel-level combination strategy surpassed the image-level and patch-level combination strategies in exploiting the complementarity of different denoised images to preserve the optimal denoised part, as well as other state-of-the-art image combination methods.

Table 4. Average performance of various combination strategies for different noise levels.

| Noise Level | $\sigma = 10$ | $\sigma = 20$ | $\sigma = 30$ | $\sigma = 40$ | $\sigma = 50$ | $\sigma = 60$ |
|-------------|---------------|---------------|---------------|---------------|---------------|---------------|
| dataset     | common10      | BSD           |               |               |               |               |
| CsNet       | 34.62         | 31.40         | 29.75         | 28.55         | 27.61         | 26.79         |
| SPD         | 34.76         | 31.37         | 29.67         | 28.41         | 27.54         | 26.77         |
| U2Fusion    | 29.66         | 28.52         | 27.63         | 26.83         | 26.24         | 25.67         |
| MEFNet      | 35.04         | 31.56         | 29.81         | 28.54         | 27.58         | 26.79         |
| Proposed    | **36.46**     | **31.81**     | **30.82**     | **28.78**     | **28.49**     | **27.03**     |

4.5. Experimental Results with Other Denoising Methods

Next, we compared the performance of the proposed model and the nine other typical denoisers. We obtained the average PSNR results of these models on 10 commonly used images with different noise levels ranging from 10 to 60 with a step of 10 and calculated the overall average at the end. From Table 5, we can obtain the following observations.
First, although CsNet combined denoised images generated by both model-based and learning-based denoisers, its average denoising effect is inferior to that of the mainstream denoisers, such as VDNet, due to its coarse combination granularity. Second, our method outperformed competing models by a considerable margin and yielded the best average PSNR results. While image denoising methods have shown encouraging achievements over the last decade, note that achieving even a slight performance increase has become more challenging for numerous denoising methods. However, the PSNR score of our method has surpassed that of the learning-based and model-based methods FFDNet and BM3D by 0.18 (on average) to 0.54 dB and 0.54 (on average) to 0.69 dB, respectively. Since only a few methods surpass the benchmark BM3D method by more than 0.3 dB on average [40,41], this is a significant improvement.

### Table 5. Average PNSR results of different denoising methods on 10 commonly used images with various noise levels.

| Noise Level | BM3D | FFDNet | NCSR | DnCNN | WNNM | VDNet | CsNet | SPD | U2Fusion | MEFNet | Proposed |
|-------------|------|--------|------|-------|------|-------|-------|-----|----------|--------|----------|
| σ = 10      | 34.81| 34.83  | 34.81| 34.94 | 34.94| 34.63 | 34.62 | 34.76| 29.66    | 35.04  | 36.46    |
| σ = 20      | 31.38| 31.68  | 31.38| 31.69 | 31.61| 31.53 | 31.40 | 31.37| 28.52    | 31.56  | 31.81    |
| σ = 30      | 29.53| 29.92  | 29.43| 29.83 | 29.80| 29.78 | 29.75 | 29.67| 27.63    | 29.81  | 30.82    |
| σ = 40      | 28.10| 28.68  | 28.06| 28.52 | 28.48| 28.37 | 28.35 | 28.41| 28.63    | 28.54  | 28.78    |
| σ = 50      | 27.47| 27.71  | 27.02| 27.56 | 27.51| 27.64 | 27.61 | 27.54| 26.24    | 27.58  | 28.49    |
| σ = 60      | 26.33| 26.90  | 26.08| 26.65 | 26.68| 26.79 | 26.77 | 26.67| 25.67    | 26.79  | 27.03    |
| Average     | 29.60| 29.96  | 29.86| 29.84 | 29.84| 29.79 | 29.75 | 27.43| 29.89    | 30.14  |          |

To demonstrate the robustness of our method, we performed comparisons by randomly selecting 50 images from the BSD, making the process of image denoising more challenging. The quantitative comparison with the competitors on the BSD is summarized in Table 6. On this dataset, the denoising effects of each denoiser show a decline to different degrees compared to the average values obtained on the first dataset. Nevertheless, it is also apparent that our method still achieved a substantial improvement over others. The PSNR results of the ten denoisers are shown in Table 6. Specifically, we observed under the average value that the PSNR score of our method surpassed that of BM3D by 0.83 dB, FFDNet by 0.30 dB, NCSR by 0.66 dB, DnCNN by 0.27 dB, WNNM by 0.43 dB, VDNet by 0.19 dB, CsNet by 0.29 dB, SPD by 0.28 dB, U2Fusion by 3.71 dB and MEFNet by 0.25 dB. Therefore, our proposed method brings significant improvements to the PSNR results.

### Table 6. Average performance of various methods on BSD with different noise levels.

| Noise Level | BM3D | FFDNet | NCSR | DnCNN | WNNM | VDNet | CsNet | SPD | U2Fusion | MEFNet | Proposed |
|-------------|------|--------|------|-------|------|-------|-------|-----|----------|--------|----------|
| σ = 10      | 33.25| 33.56  | 33.42| 33.69 | 33.53| 33.43 | 33.59 | 33.57| 28.47    | 34.19  | 34.22    |
| σ = 20      | 29.14| 29.65  | 29.58| 29.70 | 29.73| 29.93 | 29.67 | 29.59| 27.14    | 30.03  | 30.82    |
| σ = 30      | 27.34| 27.88  | 27.60| 27.87 | 27.81| 27.89 | 27.60 | 27.59| 26.14    | 28.06  | 28.16    |
| σ = 40      | 26.11| 26.73  | 26.27| 26.87 | 26.54| 26.67 | 26.75 | 26.60| 25.36    | 26.80  | 26.91    |
| σ = 50      | 25.28| 25.89  | 25.35| 25.85 | 25.63| 25.99 | 25.91 | 25.80| 24.75    | 25.92  | 26.04    |
| σ = 60      | 24.30| 25.23  | 24.59| 25.13 | 24.92| 25.31 | 25.20 | 25.13| 24.25    | 25.25  | 25.35    |
| Average     | 27.63| 28.16  | 27.80| 28.19 | 28.03| 28.27 | 28.17 | 28.08| 26.02    | 28.38  | 28.46    |

The experimental results showed that our method outperformed the other state-of-the-art denoisers on two representative datasets. In particular, our method yields a notably superior denoising effect to any other stand-alone denoising method since it makes full use of both internal and external information. As a pixel-level combination denoiser, it employs two complementary denoisers to boost the denoising effect, which is better than that of CsNet, U2Fusion, MEFNet and SPD under various noise levels in different datasets. In the following section, a visual comparison of images denoised by typical methods is presented to further support our conclusions.
4.6. Visual Comparisons

In image processing, visual quality is an important criterion for evaluating denoising effects [42]. To intuitively perceive the visual quality, we present the comparison results of images with rich texture information. Figure 7 visually shows the denoising results of the Couple image from common10 with a noise level of 30 using different denoising methods. We chose the portion of the Couple image with curtains, which was enlarged in the lower right corner of each image for a detailed comparison. It can be observed that BM3D is not effective since a lot of detail is lost in the location indicated by the red arrow on the right, and, compared to the original image, the portion of the curtain texture indicated by the yellow arrow is also distorted. NCSR cannot preserve the details of the curtain folds and its denoised image is severely distorted, especially in the right part. For the denoised image by VDNet, the blurry spots from the noise cannot be eliminated effectively, resulting in an undesirable result. CsNet and WNNM produce smooth edges; however, the texture details are not adequately preserved. While FFDNet and DnCNN preserve more texture details, they are more likely to produce over-smoothened artifacts. In the case of the proposed method, we observed from the enlarged part of the image that the details of the curtain folds were improved, and the discernibility of the curtain edge was enhanced, which is much closer to the original image. Overall, our method yielded a satisfactory denoised image with better visual effects and increased the PSNR value to 29.43 dB when compared with the aforementioned seven methods.

![Figure 7. Cont.](image-url)
Figure 7. Denoising results of Couple image with noise level $\sigma = 30$.

The visual results of the typical methods on the Hill image are shown in Figure 8. We magnified the roof part of the image to compare the visual differences in detail. BM3D cannot effectively handle the boundary part of the roof, making the edges blurred. The denoising results of FFDNet and VDNet, which are representatives of the learning-based methods, are too smooth and completely lose the details of the roof section. The NCSR and DnCNN preserve a small portion of the roof texture but it is still not apparent. Moreover, compared to the original image, there are distortions and artifacts, indicated by the arrows. As a representative of the image combination strategy, CsNet does not show superior results in this region compared to other stand-alone denoisers. This is because the granularity of its image combination strategy is too large, making it difficult to obtain optimal denoised results in some regions of the images. The performance in the magnified region of our method is superior to that of the six above methods. Specifically, our method achieves competitive results in terms of the edges of the roof details, and the textural regions are effectively retained, preserving information in the original image to the greatest extent. Since our denoising method is extensible, this allows us to easily extend the denoising method to color image denoising. Therefore, we conducted experiments with color versions of the common dataset as well.

Figure 8. Cont.
Figure 8. Denoising results of Hill image with noise level $\sigma = 30$.

For color image denoising, we chose a color version of the Boat image and the results are shown in Figure 9. We magnified the roof part of the image to compare the visual differences in detail. Since the typical BM3D method can only process grayscale images, we chose its variant, CBM3D, to conduct the visual comparison experiments on color images. In the denoised image of FFDNet, we observed that the detail part of the white line on the boat, indicated by the yellow arrow, was preserved well, but the lower right corner of the magnified area, indicated by the red arrow, showed a serious distortion, where many details were ignored. In contrast, CBM3D processed the image well at the white line indicated by the red arrow but failed to preserve the details indicated by the yellow arrow. Our method takes full advantage of the combined features of CBM3D and FFDNet, resulting in optimal denoised results. Based on the denoised image, the proposed method is visually superior to other methods. Since the combination of BM3D and FFDNet makes full use of both internal and external information, the high-quality regions of both CBM3D and FFDNet were combined into our image and it yielded a superior denoising effect.
4.7. Low-Light Image Enhancement

To demonstrate the flexibility and extensibility of our method, we applied it to low-light image enhancement and conducted a comparison experiment. Similar to our image combination denoising process, in this experiment, the low-light image is equivalent to the noisy image, and the enhanced image after processing by image enhancement methods is equivalent to the initial denoised image. Using the proposed UWMGN, we set a weight map for each initial enhanced image, and then combined the initial enhanced images to obtain the final combined enhanced image. We used the information content weighted peak signal to noise ratio (IW-PSNR), which is an extension of PSNR and is more suitable for measuring the quality of low-light enhanced images [43], as our image quality evaluation metric to measure the enhancement effect of various methods. The larger the value of IW-PSNR, the better the image enhancement effect. We used four representative image enhancement methods, viz. adaptive gamma correction with weighted distribution (AGCWD) [44], fusion-based enhancing method (Fu) [45], a new image contrast enhancement algorithm proposed by Ying et al. (Ying) [46] and a structure-revealing low-light image enhancement method (SRLLEI) [47], as a comparison. Our method combines the three image enhancement methods, viz. Fu, Ying and SRLLEI. Figure 10 shows the enhancement effect of each method on the typical low-light image selected from [48], with one area enlarged for a clearer view. We can observe from the listed IW-PSNR values that our method ranks first. In terms of visual comparison, our method produces the optimal overall performance in terms of detail protection, color fidelity and avoidance of overexposure.
Because of the selection of several complementary images to be combined, the optimal texture details of each image are retained in the final combined image.

Figure 10. Enhancement effect of each method on the typical low-light image selected from [48].

5. Conclusions

In this study, a weight map generative network was proposed. Rather than designing a novel denoising model from scratch, we combined the initial denoised images processed by different typical and effective denoisers at the pixel level to produce the optimal denoised images. Specifically, for the experiment, we found that the combination of BM3D and FFDNet has the best performance. For BM3D, we used the non-local self-similarity such that the internal information of the target image can be fully mined, alleviating the excessive reliance on the dataset. For FFDNet, we utilized its deep learning network to construct an external prior. After the two models completed the denoising process for the noisy image separately, the superior part of the two results was combined to generate an optimal denoising result via corresponding weight maps. We used an unsupervised deep learning approach to generate the weight map, as manually setting a specific weight value for each pixel point of an image is complex and difficult. This method is called the pixel-level image combination strategy. The experimental results demonstrate that our method presents a significant overall improvement compared to the results of other state-of-the-art methods, surpassing the second-best method by 0.03 dB to 1.42 dB on average, on two datasets with different noise levels. In particular, our method increased the PSNR results by an average of 0.69 dB compared to BM3D, 0.48 dB compared to FFDNet and 0.34 dB compared to CsNet. The method also performed well in visual comparisons in terms of image details and textures, in processing both common and color test images. It should be noted that our method is an extensible image combination strategy, which is reflected in two aspects. Our method is flexible in the selection of the type and number of initial denoising methods. As new denoising methods have been continuously proposed, future studies can select more efficient complementary methods to further increase the performance in terms of execution efficiency and denoising effect. Our method is based on a general image combination strategy, indicating that it is not only limited to image denoising. It would be reasonable to extend our image combination strategy to other image processing problems, such as image deblurring or low-light image enhancement.
Generally speaking, the metrics for evaluating denoising results are mainly in two aspects: effect and efficiency. For every new denoising method, the most important goal is to improve the denoising effect. In this work, we focus on a combination of stand-alone methods to improve the denoising effect; therefore, the processing time is inevitably longer than that of stand-alone denoising methods, and investigating how to reduce the processing time was beyond the scope of this study. In future research, we will explore methods to reduce the processing time while maintaining an excellent denoising effect. Specifically, we may achieve this in the following ways: with the development of denoising techniques, a natural but fundamental idea is to find some more efficient stand-alone denoising methods to reduce the pre-processing time; moreover, we can also improve the efficiency of our method by optimizing our weight map generative network. At present, there are two general implementations of combination networks, unsupervised and supervised. Our network belongs to the unsupervised approach, which allows its network parameters to be adapted to any given initial denoised images without requiring training pairs. This image-specific denoising network allows us to produce optimal combined images but with a longer processing time. Although supervised networks have advantages in terms of testing time, they require a large number of labeled training images, which are expensive to collect. In addition, most supervised combination networks suffer from poor generalization ability and fail to achieve optimal performance on testing images that are different from the training images. The recently proposed U2Fusion [37] has inspired us to develop a novel method to combine the advantages of both supervised and unsupervised approaches. We can design our proposed network to cooperate with a supervised network such as DenseNet [37] in U2Fusion. For a given noisy image, our UWMGN always produces its optimal denoising image. Therefore, in the case of no ground truth images, the labeled images required by the supervised network in the training phase will be generated by our UWMGN. As such, in the testing phase, the combined images will be generated directly from the trained supervised network, which improves the execution efficiency while maintaining excellent denoising results.

**Author Contributions:** Conceptualization, L.Y. and S.X.; methodology, S.X.; software, X.C.; validation, L.Y., J.L., S.X. and N.X.; formal analysis, L.Y. and S.X.; investigation, L.Y. and J.L.; resources, S.X.; data curation, L.Y. and X.C.; writing—original draft preparation, L.Y.; writing—review and editing, L.Y., J.L., S.X. and N.X.; visualization, L.Y. and S.X.; supervision, S.X.; project administration, S.X.; funding acquisition, S.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Natural Science Foundation of China, grant numbers 62162043, 62162042 and 62102174.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** The data presented in this study are openly available in Github (https://github.com/NCU-YLJ/UWMGN).

**Acknowledgments:** The authors would like to thank the anonymous reviewers and AE for their constructive comments and suggestions, which improved the quality of the paper.

**Conflicts of Interest:** The authors declare no conflicts of interest.

**References**

1. Healey, G.E.; Kondepudy, R. Radiometric CCD camera calibration and noise estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* 1994, 16, 267–276. [CrossRef]
2. Zhang, K.; Zuo, W.; Gu, S.; Zhang, L. Learning deep CNN denoiser prior for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3929–3938.
3. Pandey, B.K.; Pandey, D.; Wariya, S.; Agarwal, G. A deep neural network-based approach for extracting textual images from deteriorate images. *EAI Endorsed Trans. Ind. Netw. Intell. Syst.* 2021, 8, e3. [CrossRef]
4. Ignatov, A.; Byeon-Su, K.; Timofte, R.; Pouget, A. Fast camera image denoising on mobile gpus with deep learning, mobile ai 2021 challenge: Report. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2515–2524.

5. 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]

6. Fan, C.M.; Liu, T.J.; Liu, K.H. SUNet: Swin transformer UNet for image denoising. arXiv 2022, arXiv:2202.14009.

7. Liu, Y.; Wang, Z.; Zeng, Y.; Zeng, H.; Zhao, D. PD-GAN: Perceptual-details GAN for extremely noisy low light image enhancement. In Proceedings of the ICASSP 2021–2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Toronto, ON, Canada, 6–11 June 2021; pp. 1840–1844.

8. Chen, S.; Xu, S.; Chen, X.; Li, F. Image denoising using a novel deep generative network with multiple target images and adaptive termination condition. Appl. Sci. 2021, 11, 4803. [CrossRef]

9. Buades, A.; Coll, B.; Morel, J.M. A non-local algorithm for image denoising. In Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05), San Diego, CA, USA, 20–25 June 2005; Volume 2, pp. 60–65.

10. Mosseri, I.; Zontak, M.; Irani, M. Combining the power of internal and external denoising. In Proceedings of the IEEE international conference on computational photography (ICCP), Cambridge, MA, USA, 19–21 April 2013; pp. 1–9.

11. Shreyamsha Kumar, B. Image denoising based on non-local means filter and its method noise thresholding. Signal Image Video Process. 2013, 7, 1211–1227. [CrossRef]

12. Babov, K.; Foi, A.; Katkovnik, V.; Egiazarian, K. Image restoration by sparse 3D transform-domain collaborative filtering. In Image Processing: Algorithms and Systems VI; SPIE: Bellingham, WA, USA, 2008; Volume 6812, pp. 62–73.

13. Sharma, A.; Chaurasia, V. MRI denoising using advanced NLM filtering with non-subsampled shearlet transform. Signal Image Video Process. 2021, 15, 1331–1339. [CrossRef]

14. Dong, W.; Zhang, L.; Shi, G.; Li, X. Nonlocally centralized sparse representation for image restoration. IEEE Trans. Image Process. 2012, 22, 1620–1630. [CrossRef]

15. Ren, C.; He, X.; Wang, C.; Zhao, Z. Adaptive consistency prior based deep network for image denoising. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 8596–8606.

16. Wu, C.; Gao, T. Image denoise methods based on deep learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2862–2869. [CrossRef]

17. 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]

18. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.

19. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In International Conference on Machine Learning; PMLR: New York, NY, USA, 2015; pp. 448–456.

20. 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]

21. Guo, S.; Yan, Z.; Zhang, K.; Zuo, W.; Zhang, L. Toward convolutional blind denoising of real photographs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1712–1722.

22. Yue, Z.; Yong, H.; Zhao, Q.; Meng, D.; Zhang, L. Variational denoising network: Toward blind noise modeling and removal. arXiv 2019, arXiv:1908.11314.

23. Liang, Q.; Cassayre, F.; Owsianko, H.; Helou, M.E.; Süsstrunk, S. Image denoising with control over deep network hallucination. arXiv 2022, arXiv:2201.00429.

24. Tian, C.; Fei, L.; Zheng, W.; Xu, Y.; Zuo, W.; Lin, C.W. Deep learning on image denoising: An overview. Neural Netw. 2020, 131, 251–275. [CrossRef]

25. Huang, T.; Li, S.; Jia, X.; Lu, H.; Liu, J. Neighbor2neighbor: Self-supervised denoising from single noisy images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14781–14790.

26. Tian, C.; Xu, Y.; Zuo, W. Image denoising using deep CNN with batch renormalization. Neural Netw. 2020, 121, 461–473. [CrossRef] [PubMed]

27. Gu, S.; Zhang, L.; Zuo, W.; Feng, X. Weighted nuclear norm minimization with application to image denoising. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 2862–2869.

28. Choi, J.H.; Elgendy, O.A.; Chan, S.H. Optimal combination of image denoisers. IEEE Trans. Image Process. 2019, 28, 4016–4031. [CrossRef] [PubMed]

29. Cui, J.; Gong, K.; Guo, N.; Wu, C.; Meng, X.; Kim, K.; Zheng, K.; Wu, Z.; Fu, L.; Xu, B.; et al. PET image denoising using unsupervised deep learning. Eur. J. Nuclear Med. Mol. Imaging 2019, 46, 2780–2789. [CrossRef]

30. Jaggi, M. Convex optimization without projection steps. arXiv 2011, arXiv:1108.1170.

31. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. 2004, 13, 600–612. [CrossRef]
33. Krull, A.; Vičar, T.; Prakash, M.; Lalit, M.; Jug, F. Probabilistic noise2void: Unsupervised content-aware denoising. *Front. Comput. Sci.* 2020, 2, 5. [CrossRef]

34. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 234–241.

35. Rudin, L.I.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D Nonlinear Phenom.* 1992, 60, 259–268. [CrossRef]

36. Gao, X.; Lu, W.; Tao, D.; Li, X. Image quality assessment based on multiscale geometric analysis. *IEEE Trans. Image Process.* 2009, 18, 1409–1423. [PubMed]

37. Xu, H.; Ma, J.; Jiang, J.; Guo, X.; Ling, H. U2Fusion: A unified unsupervised image fusion network. *IEEE Trans. Pattern Anal. Mach. Intell.* 2020, 44, 502–518. [CrossRef] [PubMed]

38. Ma, K.; Duannu, Z.; Zhu, H.; Fang, Y.; Wang, Z. Deep guided learning for fast multi-exposure image fusion. *IEEE Trans. Image Process.* 2019, 29, 2808–2819. [CrossRef] [PubMed]

39. Arbelaez, P.; Maire, M.; Fowlkes, C.; Malik, J. Contour detection and hierarchical image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2010, 33, 898–916. [CrossRef]

40. Levin, A.; Nadler, B. Natural image denoising: Optimality and inherent bounds. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 2833–2840.

41. Levin, A.; Nadler, B.; Durand, F.; Freeman, W.T. Patch complexity, finite pixel correlations and optimal denoising. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 73–86.

42. Lin, W.; Kuo, C.C.J. Perceptual visual quality metrics: A survey. *J. Vis. Commun. Image Represent.* 2011, 22, 297–312. [CrossRef]

43. Wang, Z.; Li, Q. Information content weighting for perceptual image quality assessment. *IEEE Trans. Image Process.* 2010, 20, 1185–1198. [CrossRef]

44. Rani, S.; Kumar, M. Contrast enhancement using improved adaptive gamma correction with weighting distribution technique. *Int. J. Comput. Appl.* 2014, 101. [CrossRef]

45. Fu, X.; Zeng, D.; Huang, Y.; Liao, Y.; Ding, X.; Paisley, J. A fusion-based enhancing method for weakly illuminated images. *Signal Process.* 2016, 129, 82–96. [CrossRef]

46. Ying, Z.; Li, G.; Ren, Y.; Wang, R.; Wang, W. A new image contrast enhancement algorithm using exposure fusion framework. In *International Conference on Computer Analysis of Images and Patterns*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 36–46.

47. Wu, Y.; Song, W.; Zheng, J.; Liu, F. Noisy low-Light image enhancement using reflectance similarity prior. In Proceedings of the 2020 15th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 6–9 December 2020; Volume 1, pp. 160–164.

48. Cai, J.; Gu, S.; Zhang, L. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Trans. Image Process.* 2018, 27, 2049–2062. [CrossRef]