IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 22, 2021, accepted March 25, 2021, date of publication April 5, 2021, date of current version April 13, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3070809*

# Efficient Attention Fusion Network in Wavelet Domain for Demoireing

**CHUNYUN SUN, HUICHENG LAI, LIEJUN WANG, AND ZHENGHONG JIA**
College of Information Science and Engineering, Xinjiang University, Urumqi 830046, China
Key Laboratory of Signal Detection and Processing, Xinjiang University, Urumqi 830046, China

Corresponding author: Huicheng Lai (lai@xju.edu.cn)

This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant U1903213 and Grant U1803261, and in part by the Science and Technology Project on aid to Xinjiang Uygur Autonomous Region through the College of Software, Xinjiang University, Urumqi, China, under Grant 2019E0215.

**ABSTRACT** When taking pictures of electronic screens or objects with high-frequency textures, people often run across colorful rainbow patterns that are known as ''moire'', seriously affecting the image quality and subsequent processing. Current methods for removing moire patterns mostly extract multiscale information by downsampling pooling layers, which may inevitably cause information loss. To address this issue, this paper proposes a demoireing method in the wavelet domain. By employing both discrete wavelet transform (DWT) and inverse discrete wavelet transform (IDWT) instead of traditional downsampling and upsampling, this method can effectively increase the network receptive field without information loss. In addition, to further reconstruct more details of moire patterns, this paper proposes an efficient attention fusion module (EAFM). With a combination of efficient channel attention, spatial attention and local residual learning, this module can self-adaptively learn various weights of feature information at different levels and inspire the network to focus more on effective information such as moire details to improve learning and demoireing performance. Extensive experiments based on public datasets have shown that this suggested method can efficiently remove moire patterns and has a good quantitative and qualitative performance.

**INDEX TERMS** Demoire, deep learning, wavelet transform, attention mechanism.

## I. INTRODUCTION

Moire patterns have very important value for studying and applying in many fields, such as measurement and analysis [1] and image detection [2]. However, the moire pattern in natural images may seriously affect image quality and follow-up processing. While taking photos of electronic screens or objects with high-frequency textures, a moire pattern will inevitably be produced [3]. With the development of digital imaging technology and the popularity of digital cameras and digital screens, there is an increasing number of moire images. In recent years, demoire technology has gained increasing attention, there is a great demand for the post-processing technology of moire pattern removal.

There are two methods for generating moire images in real life. First, moire images can be produced due to aliasing caused by insufficient sampling of fine regular patterns in natural scenes. For example, when photographing knitted fabrics with fine structures or long-distance buildings, a moire pattern often appears, which mostly includes high-frequency information centered in a certain area of the image, also known as a ''texture moire image''. Second, because of the interference between the pixel grid of camera sensors and digital screens during photographing, a moire pattern can be generated and regarded as a manifestation of the phenomenon of beats, which is called a ''screen moire image'', spanning a wide frequency range to cover the whole image and more often occurs in our daily life.

Moire patterns are complex and changeable and spatially characterized by various stripes, curves or ripples. It can result in color changes and is sensitive to slight displacement variations. Therefore, by altering the shooting direction or distance, different moire patterns can be produced. As a consequence, it is difficult to determine moire distribution and thus challenging demoire.

The classical method of demoireing adds a low-pass filter in front of the photosensitive components [4] or uses an interpolation algorithm in the output of a color filter array

The associate editor coordinating the review of this manuscript and approving it for publication was Naveed Ur Rehman.

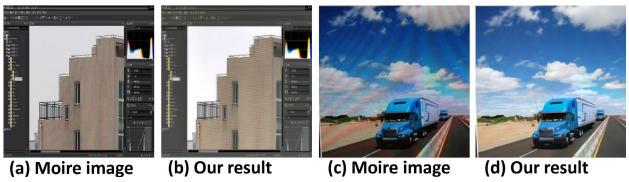


**FIGURE 1.** Moire images and our demoire results.

(CFA) [5], but it makes the image smooth, the computational complexity is high, and the actual effect is not good. Methods based on filtering or signal decomposition [6], [7] to remove moieties are mainly for high-frequency moieties in texture images, which cannot process moieties across low-frequency and high-frequency screen images at the same time. With the outstanding performance of deep learning in computer vision, deep learning-based methods have been widely applied in demoireing. Sun *et al.* [8] proposed a multiresolution (multiscale) convolutional neural network (MSCNN) for demoireing, creating a moire dataset founded on ImageNet [9]. Liu *et al.* [10] suggested a new method for demoireing composite screened moire images via a multiscale neural network from coarse to fine. In this way, moire patterns could be removed at different scales, and better visual effects can be achieved referring to generative adversarial networks (GANs) [11]. Academic have also noted this issue and held related competitions to promote research on image demoireing [12]–[18]. Existing methods for demoireing usually depend on downsampling the pooling layer to introduce a multiscale strategy, which may bring out image information loss even with good removal performance.

Inspired by MWCNN [19], to address moire patterns in screened images, this paper proposes a multiscale demoire network in the wavelet domain based on efficient attention fusion. By replacing the pooling and unpooling in the traditional UNet network with discrete wavelet transform (DWT) and inverse discrete wavelet transform (IDWT), down- and upsampling can be carried out. Due to the reversibility of DWT, the image information loss caused by up- and downsampling can thus be avoided with the ensured large receptive field of the network. Moreover, to address the uneven distribution of moire patterns, this paper introduced an efficient attention fusion module (EAFM) made up of efficient channel attention, spatial attention and local residual learning. By adjusting the output proportion self-adaptively according to input information, this module can give different weights to features at different levels to improve the representation power of the network as well as reconstructed image qualities. Fig. 1 shows the moire images and our demoire results. In summary, the main contributions are as follows:

- We propose a novel demoireing network in wavelet domain. By embedding the DWT and IDWT into Unet to replace traditional downsampling and upsampling, proposed method can effectively enlarge the receptive field without information loss, and get extra high-frequency information to improve the performance of the network.
- We propose an efficient attention fusion module (EAFM) to further focus more on moire details and adaptively learn the different weights of different feature information.
- We implement perceptual loss to enhance the visual consistency between the output image and the clean image.
- Extensive experiments on benchmark dataset demonstrate that our method outperform the existing state-of-the-art methods.

## II. RELATED WORK

In this section, we first briefly introduce the relevant methods of image demoire and then introduce the related background technologies involved in this paper, including UNet, deep learning methods based on wavelets and attention mechanisms.

### A. METHODS FOR IMAGE DEMOIREING

Due to the complexity of moire patterns, image demoireing is more challenging than general image restoration. Traditional methods are mostly aimed at moire images caused by high-frequency aliasing as a result of insufficient sampling of fine regular patterns. Yang *et al.* [4] proposed eliminating moires by placing an optical low-pass filter layer in front of a phototaking lens, but it also eliminates the high-frequency information and damages the detail clarity of the image. Hazavei and Shahdoosti [5] proposed that interpolation algorithms can be used in the output of color filter arrays (CFAs), but they rely heavily on the quality of the green channel and have high computational complexity. Liu *et al.* [6] suggested an approach based on signal decomposition and guided filtering to dismiss moire patterns in texture images. However, their methods mainly focus on the moire fringes of fabrics, and the effect of dealing with low-frequency moire fringes is not good. Yang *et al.* [7] proposed a demoireing method in screened images via layer decomposition on polyphase components (LDPC). This method has high computational complexity and easily makes the image details too smooth. To remove moires, we need to deal with both low-frequency and high-frequency information, as well as the color distortion caused by moires. Traditional algorithms are too complicated to fully remove complex and changeable moire patterns.

In recent years, deep learning has achieved remarkable success in many areas, and demoireing based on convolutional neural networks has been studied extensively. Sun *et al.* [8] created a large-scale benchmark dataset TIP dataset, which founded on ImageNet [9], plays a very important role in develop demoire methods. They also proposed a nonlinear multiresolution (multiscale) convolutional neural network (MSCNN) to eliminate moiré artefacts within different frequency bands. He *et al.* [20] proposed a network consisting of a multi-scale aggregated, edge-guided, and pattern attribute-aware network to further remove the moire pattern precisely, but they need the other two pretrained network to describe the appearance properties of moire patterns.

Yang *et al.* [21] presented a new high-resolution demoire network to fully utilize relations among feature mappings of different resolutions and multi-scale information exchange to better separate and remove moire patterns, but it achieves limited success when a moiré pattern exhibits very severe large-scale coloured bands. Zheng *et al.* [22] divided image demoireing into two subquestions, namely, moire pattern removal and color restoration, and restored moire patterns by using a learnable bandpass filter and tone mapping. In the AIM 2019 Demoireing Challenge [12], Cheng *et al.* [14] proposed a multi-scale convolution network based on dynamic feature coding to remove moire dynamically. In NTIRE 2020 challenge for texture moire images, a number of excellent network architectures have been proposed [16]–[18]. Most of the existing networks use a multiscale strategy to remove moieties of different frequencies to obtain a better effect. However, maximum pooling or average pooling is usually used to obtain moire information of different scales, which inevitably leads to information loss. In this paper, wavelet transform instead of downsampling can effectively avoid this shortcoming and enhance the network expression ability.

### B. RELATED TECHNOLOGIES

#### 1) U-NET

Ronneberger *et al.* [23] presented UNet, which is a classical codec structure that was originally designed for biomedical images, but because of its excellent performance, UNet and its variants have been widely used in various subfields of computer vision. Xiaomeng *et al.* [24], by replacing each submodule of UNet with a dense connection module, proposed a full density UNet to remove artifacts in images. Ibtehaz and Rahman [25] proposed a method for combining the MutiRes module with UNet, which uses the idea of multiresolution to replace the traditional convolution layer and proposes the residual path (RES). Oktay *et al.* [26] proposed attention UNet. Before splicing the features in each resolution of the encoder with the corresponding features in the decoder, an attention module is used to readjust the output features of the encoder. This paper continues the core idea of UNet, using wavelet downsampling to obtain different scale features in the encoder, reshaping the image through inverse wavelet upsampling in the decoder, introducing a short channel to fuse the features of different scales, and enhancing the network performance to generate a more refined moire-free image.

#### 2) WAVELET-BASED DEEP-LEARNING APPROACH

A wavelet is a powerful time-frequency analysis tool with perfect reconstruction ability, and it can completely avoid any information loss during signal decomposition. The wavelet transform decomposes the image into a combination of low-frequency images and detail (high-frequency) images, which represent the different structures of the image, so it is easy to extract the structural information and detailed information of the original image. In recent years, wavelet transform has been introduced into deep learning networks [19], [27]–[31] and has achieved good results. In this paper, we mainly introduce the wavelet transform as a sampling operation method. According to Liu *et al.* [19], multilevel wavelet decomposition is combined with CNN to reduce the resolution of the feature image and improve the resolution receptive field. The wavelet is used to achieve downsampling and retain all the components, which achieves a good denoising effect. After the wavelet transform, Han and Ye [29] input high- and low-frequency subbands into different branches for separate processing and subsequent reconstruction of clear CT images. By averaging multiple components of the wavelet transform as downsampling output, Duan *et al.* [30] effectively suppressed noise and obtained SAR image segmentation with good labeling consistency. Li *et al.* [31] discussed the relationship between DWT and downsampling and achieved better image classifications by abandoning the high-frequency components of discrete wavelet transform and noise. The wavelet transform considers both spatial and frequency information, while the moire fringe overlaps with the original image, which covers a wide range in both spatial and frequency domains. So wavelet transform is very suitable for image demoire. Considering the wide distribution of moire frequencies, it is difficult to abandon a certain part of the frequency or deal with different threshold components separately. Therefore, we use the method of [19] to merge the various band components after the wavelet transform into the next CNN without losing any information, which is beneficial to image restoration.

#### 3) ATTENTION MECHANISM

In recent years, the attention mechanism has made important breakthroughs in image processing, natural language processing and other fields, which has been proven to be beneficial to improving the performance of the model. Hu *et al.* [32] proposed the classic SEnet by recalibrating the weights of channel features using the interdependencies among feature channels, which performed well in classifying images. Improving SEnet to an efficient channel attention network (ECAnet), Wang *et al.* [33] later realized cross-channel interaction without dimension reduction and thus efficiently used characteristic channel information to improve network performance. Zhang *et al.* [34] proposed a residual channel attention block (RCAB) that functioned well in image super-resolution. Woo *et al.* [35] developed a convolutional block attention module (CBAM) to combine channel and spatial information by using the average pool and maximum pool to aggregate features, dramatically increasing image classification and target detection. Qin *et al.* [36] combined pixel level, channel level attention and local residual learning and proposed a feature fusion attention network FFANet, which expanded the expression ability of CNN. On the basis of [36], this paper proposes an efficient attention module that can guide the network to address moire features and suppress unnecessary features, improve network learning efficiency and accelerate network convergence.

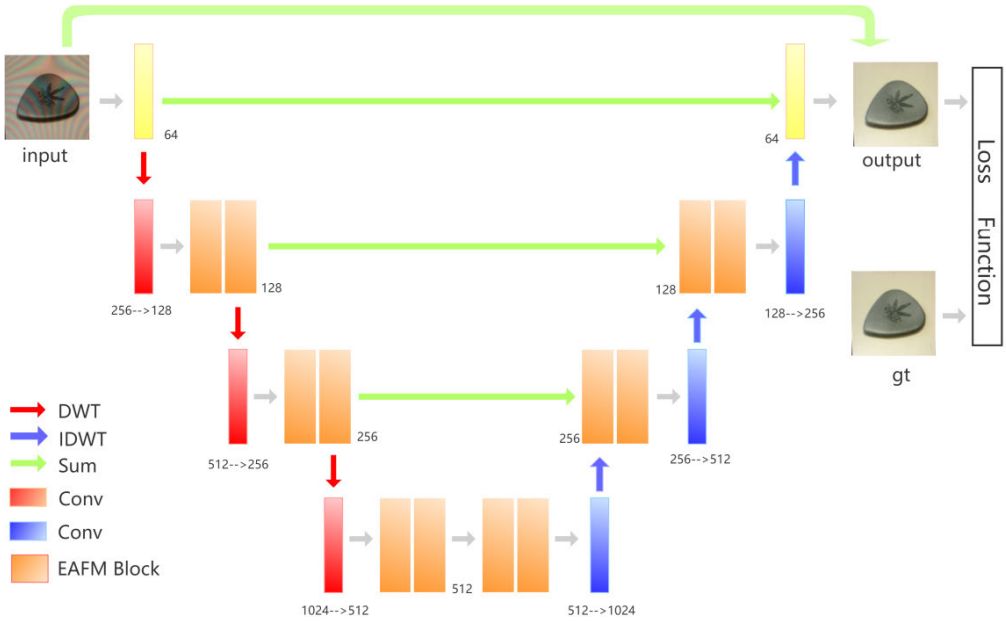


**FIGURE 2.** Network architecture of our proposed method.

## III. METHODOLOGY

In this section, we first introduce the network structure, then introduce the embedding method of wavelet transform and inverse wavelet transform, then introduce the efficient attention fusion module, and finally define the loss function optimization model.

### A. NETWORK ARCHITECTURE

The network structure is shown in Fig. 2. The input moire image $X$ is transformed from the pixel space to the feature image space by a $3 \times 3$ convolution, which changes from the RGB3 channel image to a 64-channel feature map.

$$X_i = H(X) \tag{1}$$

where $X$ indicates the input image, $H(\cdot)$ is the $3 \times 3$ convolution operation, and $X_i$ represents the shallow features obtained. Compared with the direct wavelet decomposition of the input image, the conv block is used to extract features from the input image, which is proven to be a useful image restoration by experience. Then, the shallow features obtained from spatial space are sent to a 3-level wavelet encoder-decoder module.

The wavelet encoder-decoder module is a typical structure of UNet, in which the coding part uses Haar DWT to downsample the feature map to 1/2, 1/4, and 1/8 of the original image and obtain multi-scale edge information. Then uses a convolution layer to halve the feature channel, achieve compact representation and reduce the complexity of the model. Two efficient attention fusion modules are uesd to extract features at three levels of different resolutions, capture the context information, and extract high-level semantic information. The decoding part uses the Haar IDWT to achieve upsampling. Accordingly, there is a convolution layer before IDWT to double the number of channels to achieve channel alignment. The resolution of the image is restored by the same method, reconstructed by the efficient attention fusion module. To fuse more low-level semantic information and different scale features into the final recovered feature map, a skip connection is introduced to output more detailed non-moire image features. Please note that the low-level features of the skip connection and the features used on the decoder are summed elementwise.

$$X_o = T(X_i) \tag{2}$$

where $T(\cdot)$ represents the wavelet encoder-decoder network and $X_o$ represents the reconstructed 64-channel feature map.

Finally, a $3 \times 3$ convolution layer is used to map the image from the feature space to the image space, and the reconstructed residual image is superimposed with the original image to obtain the output $Y$ of the network.

$$Y = conv(X_o) + X \tag{3}$$

### B. WAVELET TRANSFORM AND THE INVERSE

In this paper, the Haar discrete wavelet transform DWT is introduced into the network whose corresponding filter is as follows:

$$f_{LL} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \quad f_{HL} = \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix}$$
$$f_{LH} = \begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \quad f_{HH} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{4}$$

As a result, the size of the input image can be minimized, and high-frequency information can be obtained horizontally,

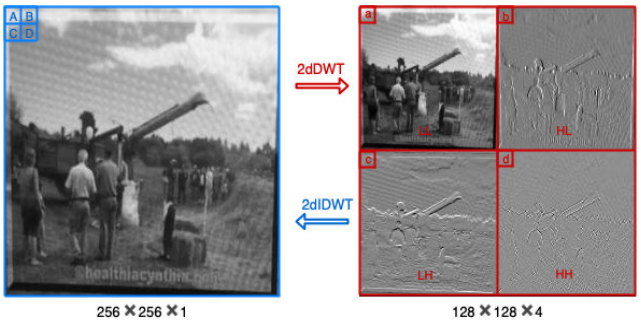


**FIGURE 3.** Downsampling of DWT-based decomposition and upsampling of IDWT-based integration of images.

vertically and diagonally, which can effectively extract edge texture details and improve network learning efficiency.

Taking the gray image as an example, as indicated in Fig. 3, A, B, C, and D and a, b, c, and d are pixel values of the corresponding images and subbands, respectively. It can be observed that DWT not only reduces the size of the input image and obtains the low-frequency information of the image $I_{LL}$, but also obtains high-frequency information in the horizontal $I_{LH}$, vertical $I_{HL}$ and diagonal directions $I_{HH}$, which is easier to extract structure and texture detail information. Given the size of an image as n × n × 1, it becomes (n/2) × (n/2) × 4 after DWT. At the same time as downsampling, the feature channel becomes 4 times the original. To reduce the number of feature mapping channels, achieve compact representation and reduce the complexity of the model, a convolution layer is used after DWT to halve the feature channel, and then it is sent to an efficient attention fusion module for feature extraction. Accordingly, a convolution layer is used before the IDWT to double the number of mapping feature channels so that each resolution feature channel is consistent. In this way, we use DWT and IDWT replace the pooling operations to reduce information loss and reserve high frequency details, enlarge the receptive field at the same time.

### C. EFFICIENT ATTENTION FUSION MODULE

Enlightened by FFA [36], this paper proposes an efficient attention fusion module (EAFM) that combines efficient channel attention (ECA), spatial attention (SA) and local residual learning (LRL), as shown in Fig. 4(a). Referring to the joint order of the attention module in CBAM [35], EAFM first scales channel features through the ECA module and then recalibrates the spatial information weight of the output features in the ECA module by the SA module according to the spatial interdependency of the input features in the module. LRL is introduced to improve discrimination learning of the module for residual information, which can better the performance and stability of the network while making the network focused more on important spatial information. The experimental results have proven that the EAFM module could effectively improve network performance.

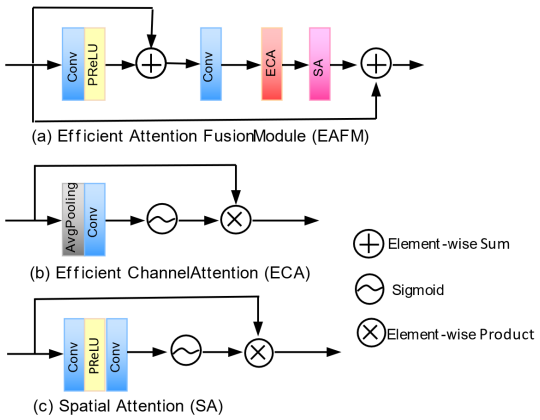


**FIGURE 4.** Efficient attention fusion module (EAFM).

#### 1) EFFICIENT CHANNEL ATTENTION (ECA)

To obtain the weights of different channels, this paper adopts efficient channel attention (ECA), as shown in Fig. 4(b). Only weights among the channel and surrounding k channels are considered in this paper, while k represents the coverage ratio of the local cross-channel interaction.

First, the global spatial information of the channel is transformed into a one-dimensional global channel tensor via global average pooling.

$$g_c = H_P(F_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} X_c(i, j) \tag{5}$$

where $X_c(i.j)$ signifies the value of $X_c$ at position $(i, j)$ in the c-th channel, and $H_P(\cdot)$ stands for the global pooling function. The feature image is converted from C×H×W to C × 1 × 1, and weight prediction of adjacent channels is realized by 1 × 1 convolution.

$$CA_c = \sigma(Conv(g_c)) \tag{6}$$

where $\sigma(\cdot)$ acts as the sigmoid function and $Conv$ is the abbreviation of a 1 × 1 convolution. Finally, by multiplying the weight of the $CA_c$ channel with the input $F_c$, a weight of the local channel interaction that relates only to the corresponding $k$ channels is obtained.

$$F_c^* = F_c \otimes CA_c \tag{7}$$

For the given channel dimension $C$, the coverage ratio k of local cross-channel interaction can be adaptively described as:

$$k = \phi(C) = \left| \frac{\log_2 C}{2} + \frac{1}{2} \right|_{odd} \tag{8}$$

where $|\cdot|_{odd}$ denotes the proximal odd. The detail information can be founded in [33].

#### 2) SPATIAL ATTENTION (SA)

Spatial attention (SA) is illustrated in Fig. 4(c). After performing PReLU and the sigmoid activation function, the input $F_c^*$ (the output of $ECA$) is directly input into two convolutional layers. As a result, the feature map changes from

**TABLE 1.** Quantitative comparisons of different methods.

| | DnCNN | U-net | DMCNN | MopNet | HRDN | Ours |
|---|---|---|---|---|---|---|
| PSNR | 24.45 | 26.49 | 26.77 | 27.75 | 28.47 | 29.7 |
| SSIM | 0.834 | 0.864 | 0.871 | 0.895 | 0.860 | 0.912 |
| FSIM | 0.901 | 0.902 | 0.914 | 0.938 | 0.926 | 0.945 |

C×H×W to 1×H×W, and the weights of each spatial space are output.

$$PA = \sigma(Conv(\delta(Conv(F_c^*)))) \tag{9}$$

Finally, element-to-element multiplication is carried out between the inputs $F_c^*$ and $SA$ to gain spatial attention, which is the output of the spatial attention module.

$$F^* = F_c^* \otimes PA \tag{10}$$

### D. LOSS FUNCTION

For a better reconstruction of images, the loss function in this paper is defined as follows:

$$L = L_{Charb} + \lambda_1 L_{percep} \tag{11}$$

where $\lambda_1$ denotes hyperparameters, $L_{Charb}$ signifies L1_Charbonnier_loss, and $L_{percep}$ denotes the perceptual loss. Assume $I$ and $\hat{I}$ are the output image predicted by the network and the ground truth; then, the following equation can be obtained:

$$L_{Charb}(I, \hat{I}) = \frac{1}{hwc} \sum_{i,j,k} \sqrt[2]{(I_{i,j,k} - \hat{I}_{i,j,k})^2 + \varepsilon^2} \tag{12}$$

where $\varepsilon = 0.001$, and the L1_Charbonnier_loss is more stable in performance and quicker in convergence than L1 and L2 losses.

To keep the output image consistent with the ground truth in subjective visual perception, a perceptual loss is introduced in this paper.

$$L_{percep}(I, \hat{I}; \varphi, l) = \frac{1}{h_l w_l c_l} \sqrt{\sum_{i,j,k} (\varphi^{(l)}_{i,j,k}(I) - \varphi^{(l)}_{i,j,k}(\hat{I}))^2} \tag{13}$$

where $\varphi^{(l)}_{i,j,k}$ denotes the feature map in the l-th layer of VGG-19 trained by ImageNet [9] and $h_l \times w_l \times c_l$ represents the size of the feature map. The features extracted at the 14th layer of VGG-19 are adopted in this paper as the input of perceptual loss.

## IV. EXPERIMENT AND DISCUSSION

### A. DATASETS AND EXPERIMENTAL DETAILS

The TIP dataset created by Sun *et al.* [8] is used for training and testing in this paper. This dataset includes 13,500 pairs of screen images with moire and corresponding ground truth, 90% of which are for training and 10% for testing. With constantly changing angle and distance, those images were captured by different mobile phones shooting against the ImageNet [9] dataset in different display screens to obtain as many kinds of moire patterns as possible.

The PyTorch framework is employed in this paper on a NVIDIA Tesla V100 GPU. And the input image is 256 × 256, the batch size is 16. The hyperparameters $\lambda_1$ is set 2,000. The learning rate starts from 0.0002 and gradually reduces when the loss function stops decreasing. After reaching 0.0001, it fails to 0.00005. With the Adam optimizer, 100 epochs are trained, and the model has 38 M parameters in total.

### B. EXPERIMENTAL RESULTS

We compare our method with three state-of-the-art demoireing methods [8], [20], [21] and moire patterns regarded as a special noise, two classical denoising methods [23], [37] are used for comparison.

Three metrics, Peak Signal-to-Noise Ratio (PSNR) [38], Structure Similarity Index (SSIM) [39] and feature similarity (FSIM) [40] are employed, larger values represent better results. To further evaluate the capability of improving image quality visually, we also implement the visual comparison with state-of arts. The results can be found in Fig. 5. The experiments suggest that this method is both simple in structure and efficient in performance.

Table 1 is the quantitative results among DncNN [37], UNet [23], DMCNN [8], MopNet [20], and HRDN [21]. Functioning as a general image denoising method, DnCNN has limited effect for demoireing. Connected by encoder-decoders with different resolutions, UNet delivers higher PSNR, SSIM and FSIM values. Although it is a particular network for demoireing MSCNN is not very efficient in improving the values. Moreover, referring to Fig. 5, it is obvious that some artifacts still remain in images after using the above three methods for demoireing. By adding moire information such as frequency, color and shape in the training, MopNet shows rather excellent results. HRDN further improves PSNR by integrating a fine high-resolution network, information exchange and feature fusion together, but the visual effects are not as good as MopNet. As seen, the sky section in the second row of Fig. 5 still has some artifacts. On the basis of HRDN, the method used in this paper increases PSNR by 1.23 dB with a better visual effect, get the highest SSIM values 0.912, it proves that our method is better than other state-of-arts in both objective and subjective measurements.

To show that the method in this paper has strong generalization, we also choose real moire images for testing, and the results are shown in Fig. 6. The first line shows moire images searched online, and the second line shows moire

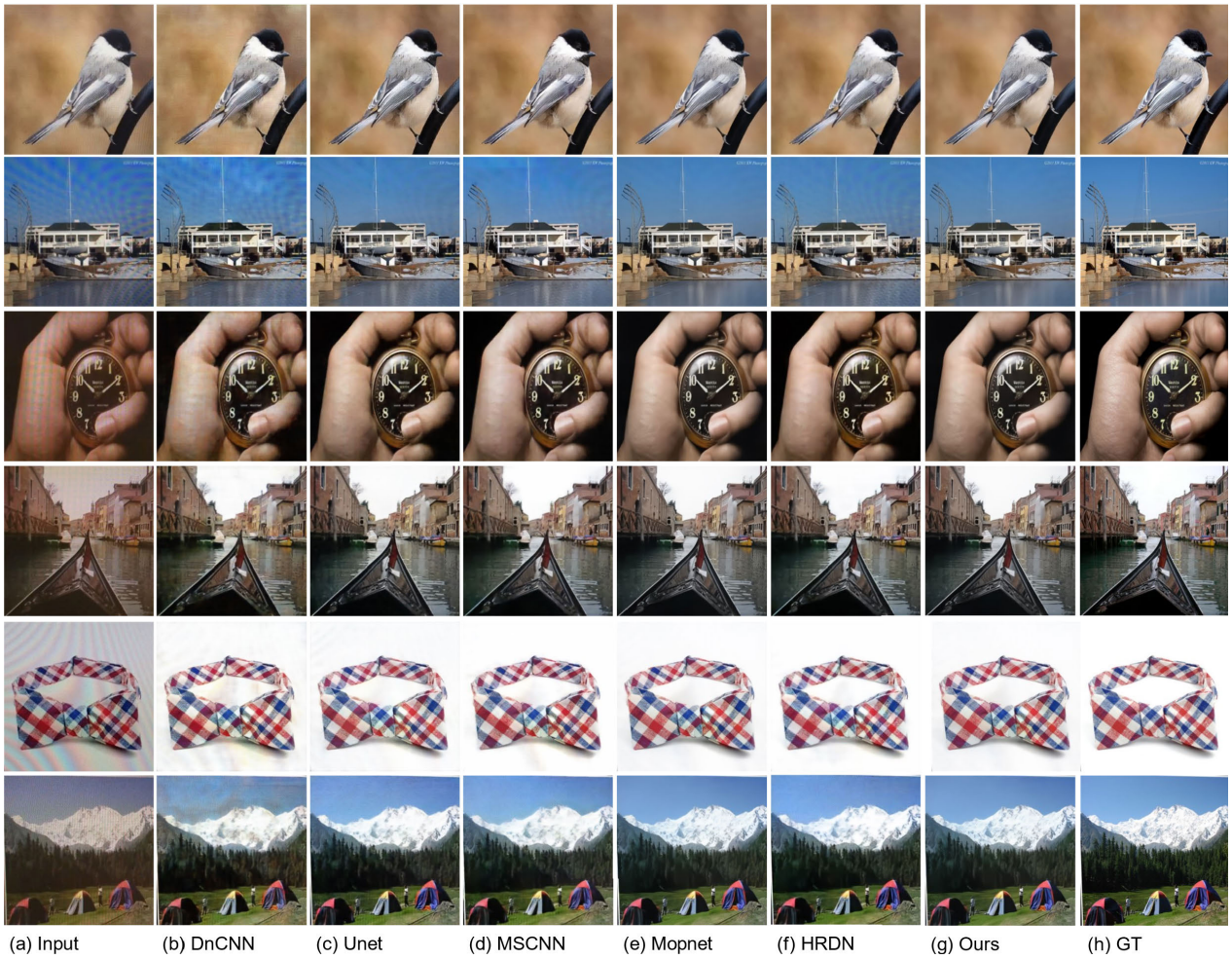


**FIGURE 5.** Qualitative comparisons of DnCNN, UNet, MSCNN, HRDN, MopNet and Ours.

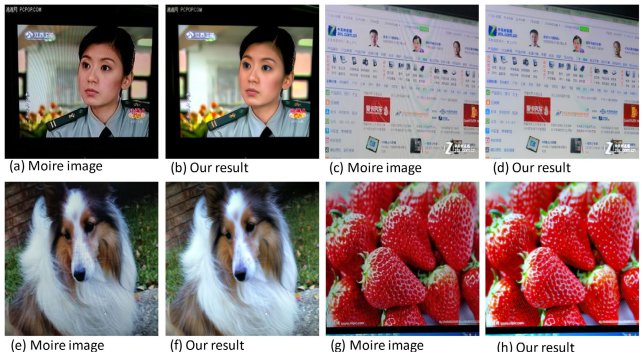


**FIGURE 6.** Real moire images and the corresponding output of our network.

images photographed by a Huawei Nova 5 Pro against a Samsung screen. As shown in Fig. 6, although the four tested images are not in our dataset, the proposed method can still effectively remove this kind of moire patterns. It shows that the dataset we use contains a variety of moire patterns, and the proposed method can grasp the details of moire patterns at various scales, so as to effectively restore the image without moire pattern.

**TABLE 2.** Ablation experiment results.

| | No DWT | No EAFM /Conv | No EAFM /RCAB | No EAFM /FFA | No percep loss | Full Ours |
|---|---|---|---|---|---|---|
| PSNR | 27.72 | 26.97 | 28.01 | 28.16 | 27.95 | 28.34 |
| SSIM | 0.876 | 0.864 | 0.879 | 0.882 | 0.869 | 0.890 |
| FSIM | 0.924 | 0.932 | 0.936 | 0.935 | 0.938 | 0.941 |

### C. ABLATION EXPERIMENT

A miniTIP dataset released by MopNet is used for ablation experiments in this paper, a simplified version of the TIP dataset. This dataset is a small dataset with approximately 1/10 of the TIP dataset volume. In this paper, 50 epochs of this dataset are trained to verify the impacts of DWT, EAFM and perceptual loss on the network structure.

The ablation experiment results are illustrated in Table 2. (1) No DWT indicates replacing DWT and IDWT with traditional pooling and unpooling in the proposed network structure, (2) no EAFM/Conv replaces the EAFM module with the ordinary convolution layer in the proposed network structure, (3) no EAFM/RCAB denotes replacing the EAFM

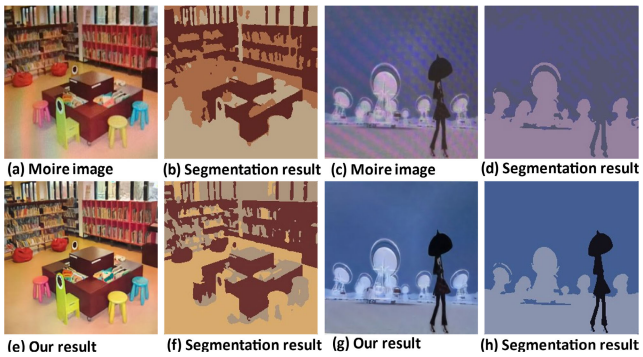


**FIGURE 7.** Segmentation results of the mean shift.

module with the RCAB [40] module in the proposed network structure, (4): no EAFM/FFA denotes using the FFA [35] module instead of the EAFM module in the proposed network structure, (5) no perception loss indicates that only L1 Charbonnier loss is used in this paper, and (6) full ours represents the proposed complete network structure. Based on points (1) and (6), as DWT and IDWT can avoid information loss during up- and downsampling, the introduction of DWT can enhance the network performance and thus increase PSNR by 0.51 dB. According to points (2), (3), (4), and (6), the network performance is improved significantly after the addition of the attention mechanism. In contrast to the RCAB and FFA modules, the RAFM module enhances PSNR by 0.33 dB and 0.18 dB respectively, with faster convergence, which can effectively improve the demoireing performance of the network. Referring to points (5) and (6), perceptual loss can not only benefit visual reconstruction but also contribute greatly to the network, which enhances PSNR by 0.39 dB.

### D. APPLICATION

It has attracted increasing attention to perform high-level computer vision tasks through data pre-processing. To investigate the impact of moire patterns on image segmentation, we use the mean shift proposed by [41] to segment moire images and our demoire images, as shown in Fig. 7.

From Fig. 7, we can see that after removing moire patterns with our method, the segmentation results get significant improvement. Especially for image Fig. 7(c), the moire pattern covers the entire image, the widespread color stripes seriously affect the visual quality of the image, so it is failed to segment the characters from the background which can be observed in Fig. 7(d). Our proposed method can remove moire patterns effectively and obtain a clear and accurate segmentation results as show in Fig. 7(h).

## V. CONCLUSION

With the advance in digital imaging technologies and the popularity of digital cameras and digital screens, moire images continue to increase in daily life. Depending on the orthogonality and reversibility of DWT, this paper proposed a new method for replacing traditional up- and downsampling with IDWT and DWT. In this respect, it can obtain all spatial resolutions with no information loss and guarantee a large receptive field of the network. In the encoder-decoder network, an efficient attention fusion module (EAFM) is adopted to extract multiscale deep features of images for multidimensional information integration, which can make the network focus more on moire details and thereby improve its performance. Additionally, L1_Charbonnier_loss and perceptual loss are also employed in network training, which improves both objective indicators and visual effects. Extensive experiments on benchmark dataset demonstrate that our method outperform the existing state-of-the-art methods.

## REFERENCES

[1] Y. Nishijima and G. Oster, "Moiré patterns: Their application to refractive index and refractive index gradient measurements," *J. Opt. Soc. Amer.*, vol. 54, no. 1, pp. 1–4, 1964.

[2] K. Patel, H. Han, A. K. Jain, and G. Ott, "Live face video vs. spoof face video: Use of Moiré patterns to detect replay video attacks," in *Proc. Int. Conf. Biometrics (ICB)*, May 2015, pp. 98–105.

[3] G. Oster and Y. Nishijima, "Moiré patterns," *Sci. Amer.*, vol. 208, no. 5, pp. 54–63, 1962.

[4] J. Yang, F. Liu, H. Yue, X. Fu, C. Hou, and F. Wu, "Textured image demoiréing via signal decomposition and guided filtering," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3528–3541, Jul. 2017.

[5] S. Mahya Hazavei and H. Reza Shahdoosti, "A new method for removing the Moire' pattern from images," 2017, *arXiv:1701.09037*. [Online]. Available: http://arxiv.org/abs/1701.09037

[6] F. Liu, J. Yang, and H. Yue, "Moiré pattern removal from texture images via low-rank and sparse matrix decomposition," in *Proc. Vis. Commun. Image Process. (VCIP)*, Dec. 2015, pp. 1–4.

[7] J. Yang, X. Zhang, C. Cai, and K. Li, "Demoiréing for screen-shot images with multi-channel layer decomposition," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.

[8] Y. Sun, Y. Yu, and W. Wang, "Moiré photo restoration using multiresolution convolutional neural networks," *IEEE Trans. Image Process.*, vol. 27, no. 8, pp. 4160–4172, Aug. 2018.

[9] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[10] B. Liu, X. Shu, and X. Wu, "Demoiréing of camera-captured screen images using deep convolutional neural network," 2018, *arXiv:1804.03809*. [Online]. Available: https://arxiv.org/abs/1804.03809

[11] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 3, 2014, pp. 2672–2680.

[12] S. Yuan, R. Timofte, G. Slabaugh, and A. Leonardis, "AIM 2019 challenge on image demoireing: Dataset and study," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3526–3533.

[13] S. Yuan, R. Timofte, G. Slabaugh, A. Leonardis, B. Zheng, X. Ye, and X. Tian, "AIM 2019 challenge on image demoireing: Methods and results," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3534–3545.

[14] X. Cheng, Z. Fu, and J. Yang, "Multi-scale dynamic feature encoding network for image demoiréing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 3486–3493.

[15] S. Yuan, R. Timofte, A. Leonardis, and G. Slabaugh, "NTIRE 2020 challenge on image demoireing: Methods and results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 460–461.

[16] X. Luo, J. Zhang, M. Hong, Y. Qu, Y. Xie, and C. Li, "Deep wavelet network with domain adaptation for single image demoiréing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 420–421.

[17] D. Xu, Y. Chu, and Q. Sun, "Moiré pattern removal via attentive fractal network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 472–473.

[18] S. Liu, C. Li, N. Nan, Z. Zong, and R. Song, "MMDM: Multi-frame and multi-scale for image demoiréing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020.
[19] P. Liu, H. Zhang, W. Lian, and W. Zuo, "Multi-level wavelet convolutional neural networks," *IEEE Access*, vol. 7, pp. 74973–74985, 2019.
[20] B. He, C. Wang, B. Shi, and L. Duan, "Mop Moiré patterns using MopNet," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 2424–2432.
[21] S. Yang, Y. Lei, S. Xiong, and W. Wang, "High resolution demoire network," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2020, pp. 888–892.
[22] B. Zheng, S. Yuan, G. Slabaugh, and A. Leonardis, "Image demoiréing with learnable bandpass filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 3636–3645.
[23] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Springer, 2015, pp. 234–241.
[24] L. Xiaomeng, C. Hao, X. Qi, D. Qi, F. Chi-Wing, and H. Pheng-Ann, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.
[25] N. Ibtehaz and M. S. Rahman, "MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.
[26] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: https://arxiv.org/abs/1804.03999
[27] T. Guo, H. S. Mousavi, T. H. Vu, and V. Monga, "Deep wavelet prediction for image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 104–113.
[28] H. Huang, R. He, Z. Sun, and T. Tan, "Wavelet-SRNet: A wavelet-based CNN for multi-scale face super resolution," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 1689–1697.
[29] Y. Han and J. C. Ye, "Framing U-Net via deep convolutional framelets: Application to sparse-view CT," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1418–1429, Jun. 2018.
[30] Y. Duan, F. Liu, L. Jiao, P. Zhao, and L. Zhang, "SAR image segmentation based on convolutional-wavelet neural network and Markov random field," *Pattern Recognit.*, vol. 64, pp. 255–267, Apr. 2017.
[31] Q. Li, L. Shen, S. Guo, and Z. Lai, "Wavelet integrated CNNs for noise-robust image classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 7245–7254.
[32] J. Hu, L. Shen, G. Sun, and S. Albanie, "Squeeze-and-excitation networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 8, pp. 2011–2023, Aug. 2020.
[33] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu, "ECA-net: Efficient channel attention for deep convolutional neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Seattle, WA, USA, 2020, pp. 11531–11539.
[34] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.
[35] S. Woo, J. Park, J. Y. Lee, and I. S. Kweon, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 3–19.
[36] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia, "FFA-Net: Feature fusion attention network for single image dehazing," *Proc. AAAI Conf. Artif. Intell.*, 2020, vol. 34, no. 7, pp. 11908–11915.
[37] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.
[38] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of PSNR in image/video quality assessment," *Electron. Lett.*, vol. 44, no. 13, pp. 800–801, Jun. 2008.
[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
[40] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: A feature similarity index for image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2378–2386, Aug. 2011.
[41] D. Comaniciu and P. Meer, "Mean shift: A robust approach toward feature space analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 5, pp. 603–619, May 2002.

**CHUNYUN SUN** received the bachelor's degree in electronic information engineering from the Communication University of China, in 2012. She is currently pursuing the master's degree in information and communication engineering with the School of Information Science and Engineering, Xinjiang University. Her current research interests include image processing and image restoration.

**HUICHENG LAI** received the B.E. and M.S. degrees from Xinjiang University, China, in 1986 and 1990, respectively. He is currently a Professor with Xinjiang University. His current research interests include image processing, image recognition, image enhancement, image restoration, and communications technology.

**LIEJUN WANG** received the Ph.D. degree from the School of Information and Communication Engineering, Xi'an Jiaotong University, in 2012. He is currently a Professor with the School of Information Science and Engineering, Xinjiang University. His research interests include wireless sensor networks, encryption algorithm, and image intelligent processing.

**ZHENGHONG JIA** received the B.S. degree from Beijing Normal University, Beijing, China, in 1985, and the M.S. and Ph.D. degrees from Shanghai Jiao Tong University, Shanghai, China, in 1987 and 1995, respectively. He is currently a Professor with the Autonomous University Key Laboratory of Signal and Information Processing, Xinjiang University, China. His research interests include digital image processing and photoelectric information detection and sensor.