Multi-spectral Dynamic Feature Encoding Network for Image Demoiréing

Qiang Dai[1], Xi Cheng[2], and Li Zhang[1(✉)]

[1] School of Computer Science and Technology,
Soochow University, Suzhou 215006, Jiangsu, China
`zhangliml@suda.edu.cn`
[2] School of Computer Science and Engineering, Nanjing University of Science
and Technology, Nanjing 210094, China
`chengx@njust.edu.cn`

Abstract. Moiré often appears when photographing textured objects, which can seriously degrade the quality of the captured photos. Due to the wide distribution of moiré and the dynamic nature of the moiré textures, it is difficult to effectively remove the moiré patterns. In this paper, we propose a multi-spectral dynamic feature encoding (MSDFE) network for image demoiréing. To solve the issue of moiré with distributed frequency spectrum, we design a multi-spectral dynamic feature encoding module to dynamically encode moiré textures. To remedy the issue of moiré textures with dynamic nature, we utilize a multi-scale network structure to process moiré images at different spatial resolutions. Extensive experimental results indicate that our proposed MSDFE outperforms the state-of-the-art in terms of fidelity and perceptual on benchmarks.

Keywords: Image demoiréing · Moiré pattern · Multi-spectral dynamic feature encoding

1 Introduction

Moiré usually appears in the form of colored stripes, ripples, or curves and is sensitive to shooting distance and camera orientation, which seriously degrades the visual quality of captured images. A common example of causing moiré is to shoot digital screens with a smartphone. More specially, digital screens are composed of liquid crystal display (LCD), which has a similar texture structure to the color filter array (CFA) of the digital camera's sensor. As a result, the imperfect alignment between them causes moiré.

Unlike other image restoration problems, such as denoising [22], super-resolution [5,15], and deblurring [12,14], a moiré pattern on images is dynamic and has a broad frequency distribution, which covers both the low-frequency part and the high-frequency parts. As a result, the broad frequency spectrum and the dynamic texture of moiré are the two main challenges in image demoiréing [3].

Recently, many methods based on convolutional neural network (CNN) [2,3,6,7,19] have been proposed to remove moiré patterns of different frequency bands. A multi-scale network architecture was typically used in these demoiréing methods to obtain dynamic representations of moiré patterns, which facilitates the modeling of moiré patterns with different spatial frequencies and different resolutions. We also utilize a multi-scale design with three branches in our proposed model, which learns the parameters of each branch and aggregates the results at different branches to produce the final output. Furthermore, a progressive upsampling strategy was employed in our model to smoothly increase the resolution.

However, existing methods cannot handle the dynamic nature of moiré textures, and none of them tried to model moiré patterns explicitly. In this paper, we propose a multi-spectral dynamic feature encoding (MSDFE) network for image demoiréing. In our model, we leverage a multi-spectral dynamic feature encoding module to encode moiré patterns, which helps the model learn the frequency prior to moiré patterns and restore moiré images clearly.

2   Proposed Method

2.1   Overall Network Architecture

As shown in Fig. 1, we construct a three-branch convolutional network to dispose of demoiréing tasks. On Branch 1, we first utilize three learnable convolutional layers to downsample and encode original moiré images, which can be described as:

$$D_1 = \bigtriangledown(I) = W_3(W_2(\delta^p W_1(I))), \tag{1}$$

where $\bigtriangledown$ denotes the downsampling block, D_1 is the output of the first downsampling block, W_i means the parameters of the i-th convolution layers in the downsampling block, δ^p is the activation function of parametric rectified linear units (PReLU) [8], and I is the input image.

Branch 1 makes use of 2 residual blocks and increases channel attention. In addition, the resolution of Branch 1 is half of the original input image. The whole process of Branch 1 can be described as follows:

$$F_1 = \triangle(MD_2(MD_1(D_1)) + D_1), \tag{2}$$

where F_1 is the output of Branch 1, $\triangle$ denotes the upsampling block with subpixel convolution [18], and MD_i means the i-th dynamic feature encoding residual block with multi-spectral channel attention, whose structure is shown in Fig. 4. We will explain the operation of multi-spectral channel attention in Sect. 2.3.

Both Branches 2 and 3 have a similar structure as Branch 1 except the number of residual blocks. The calculation processes of Branches 2 and 3 can be expressed as follows:

$$F_i = \triangle(NL(MD_k \cdots (MD_1(D_{i-1}))) + D_{i-1}), \tag{3}$$
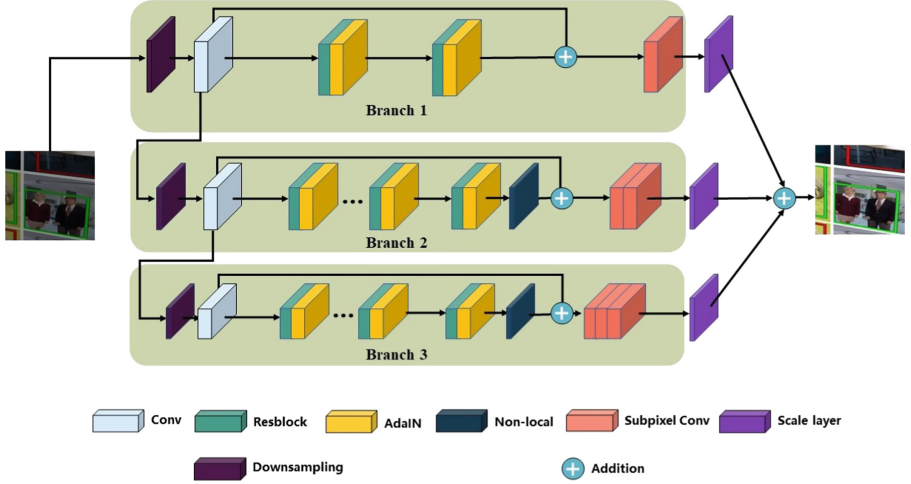
Fig. 1. Overall architecture of our proposed MSDFE.

where F_i means the output of the i-th branch, k is the number of multi-spectral dynamic feature encoding residual blocks, $NL(\cdot)$ denotes the region-level non-local operation [4] at the end of the current branch, which could help the model learn self-similarity; D_{i-1} is the downsampled feature maps from the $i-1$-th branch, whose resolution is higher.

2.2   DCT and Channel Attention

In this section, we describe DCT and channel attention.

DCT. Usually, DCT is used for data or image compression. DCT can convert the signal in the spatial domain to the frequency domain due to its strong energy concentration property. Specially, we utilize the two-dimensional DCT (2D-DCT) [1]. For an image $I \in \mathbb{R}^{H \times W}$ with a height of H and a width of W, 2D-DCT can be described as:

$$\Psi(h, w) = 2DDCT_{h,w}(I) = \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} I_{u,v} \beta_{u,v}(h, w),$$
$$h \in \{0, 1, \cdots, H-1\}, w \in \{0, 1, \cdots, W-1\}, \tag{4}$$

where $\Psi(h, w)$ denotes the 2D-DCT frequency spectrum at (h, w), $2DDCT_{h,w}$ is the 2D-DCT function, $I_{u,v}$ is the pixel of image I at (u, v), $h \in \{0, 1, \cdots, H-1\}$, $w \in \{0, 1, \cdots, W-1\}$, and the basis function $\beta_{u,v}(h, w)$ has the form

$$\beta_{u,v}(h, w) = cos\left(\frac{\pi h}{H}\left(u + \frac{1}{2}\right)\right) cos\left(\frac{\pi w}{W}\left(v + \frac{1}{2}\right)\right) \tag{5}$$
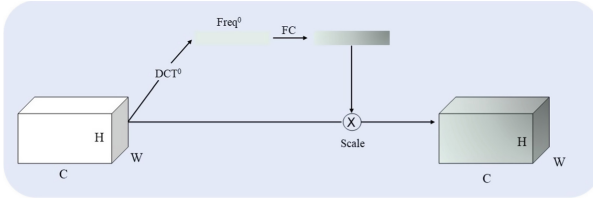
Fig. 2. Channel attention (CA).

Channel Attention. As shown in Fig. 2, the channel attention mechanism [16] uses a scalar to represent and evaluate the importance of each channel.

Let $X \in \mathbb{R}^{H \times W \times C}$ be the 2D feature maps in the network, where C means the number of channels, H and W denote the height and the width of feature maps. Qin et al. [16] found that the scalar representation in channel attention could be treated as a compression problem because it has to represent the whole channel while only one scalar can be used. From this perspective, the channel attention can be expressed as:

$$\gamma = sigmoid(fc(\Gamma_{compress}(X))), \tag{6}$$

where $\gamma \in \mathbb{R}^C$ denotes the channel attention vector, $sigmoid$ means the Sigmoid function, fc represents a mapping function, such as a fully connected layer or a 1D convolution, and $\Gamma_{compress} : \mathbb{R}^{H \times W \times C} \to \mathbb{R}^C$ is a compression method.

After obtaining the attention vector of all C channels, the attention value will scale the corresponding input X. This process could be calculated as:

$$\overline{X}'_c = \gamma_c X'_c, c \in \{0, 1, \cdots C - 1\}, \tag{7}$$

where $\overline{X}'_c$ denotes the output of attention mechanism, γ_c represents the c-th element of attention vector, and X'_c means the c-th channel of the input X.

2.3 Multi-spectral Channel Attention (MSCA)
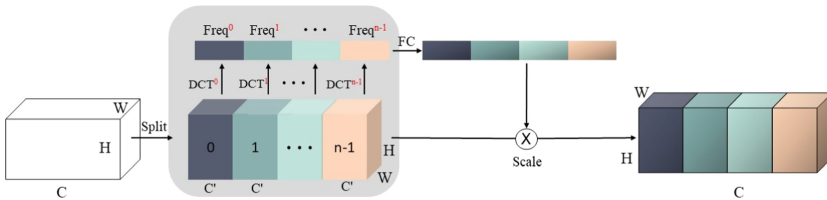
Fig. 3. Multi-spectral channel attention (MSCA [16]).

Given the input $X \in \mathbb{R}^{H \times W \times C}$, it will be divided into n parts: $[X^0, X^1, \cdots, X^{n-1}]$ along the channel dimension, where $X^j \in \mathbb{R}^{H \times W \times C'}, C' =$

$\frac{C}{n}$. For each part, a corresponding 2D-DCT frequency component is assigned. This can be formulated as:

$$Freq^j = \Gamma_{compress}(2DDCT_{h_j,w_j}(X^j)), \tag{8}$$

where $[h_j, w_j]$ are the frequency component 2D indices corresponding to X^j, $Freq^j \in \mathbb{R}^{C'}$ is the C' dimensional vector after the compression. More especially, the compression vector could be acquired by concatenating all the dimensional vectors, which can be calculated as:

$$Freq = concat([Freq^0, Freq^1, \cdots, Freq^{n-1}]), \tag{9}$$

where $Freq \in \mathbb{R}^C$ denotes the multi-spectral vector. In fact, the MSCA can be described as [16]:

$$MSCA = sigmoid(fc(Freq)), \tag{10}$$

The channel information after compression is effectively enriched for representation, and the overall illustration of the MSCA is shown in Fig. 3.

2.4   Multi-spectral Dynamic Feature Encoding (MSDFE)

We encode image feature maps at different spatial resolutions using three branches with various scales. We also add MSCA [16] and adaptive instance normalization (AdaIN) [10] in each branch. In addition, the residual block [9] is used to model the difference between clean and moiré images at each feature level and frequency band.

Afterwards, we calculate the mean value and the variance of the obtained MSCA feature maps in the i-th encoding layer as:

$$\hat{\mu}_i = \frac{1}{HW} \sum_{u=1}^{H} \sum_{v=1}^{W} X_{uv}^i, \tag{11}$$

and

$$\hat{\sigma}_i^2 = \frac{1}{HW} \sum_{u=1}^{H} \sum_{v=1}^{W} (X_{uv}^i - \mu_i)^2, \tag{12}$$

where H and W denote the height and width of the MSCA feature map, $\hat{\mu}_i$ and $\hat{\sigma}_i^2$ are the mean value and the variance of the MSCA feature map from the i-th encoding layer in the multi-spectral dynamic feature encoding branch, respectively.

We utilize statistical values that we calculate from moiré patterns to dynamically adjust the statistical characteristics (i.e. mean and variance) of each branch via AdaIN as following:

$$X^{i+1} = \frac{X^i - \mu_i}{\sqrt{\sigma_i^2 + \epsilon}} \sqrt{\hat{\sigma}_i^2} + \hat{\mu}_i, \tag{13}$$

where X^i denotes the feature maps from the i-th residual block in each branch, μ_i and σ_i^2 mean the statistical information from the multi-spectral dynamic feature encoding branch.
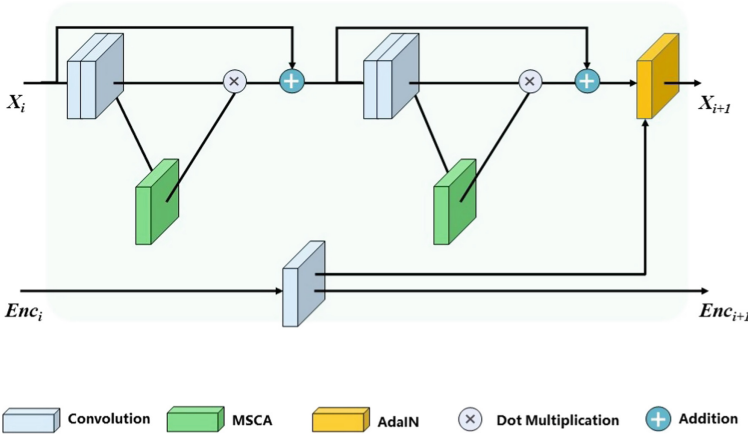
Fig. 4. Multi-spectral channel attention dynamic feature encoding residual block, where MSCA denotes the multi-spectral channel attention.

2.5   Loss Function

In terms of the loss function, Lai et al. [13] found that using the general L_2 loss in image restoration would make the image over smooth. To make demoiréd images more visually pleasing, we refer to the Charbonnier loss proposed by Lai et al., which can be calculated as the following:

$$Loss_{char} = \frac{1}{B} \sum_{i=1}^{B} \sqrt{(I_i - \hat{I}_i)^2 + \varepsilon^2}, \tag{14}$$

where B denotes the batch size of the input, ε is a parameter in the Charbonnier penalty and is empirically set to 0.001, and I_i and $\hat{I}_i$ are the ground truth and generated demoiréd image, respectively.

Besides, we also utilize the L_1 wavelet loss proposed by Cheng et al. [3] to address moiré patterns with different spatial frequencies, which has the form:

$$Loss_{wav} = \frac{1}{B} \sum_{i=1}^{B} \sqrt{\left(\Gamma_{wav}(\hat{I}_i) - \Gamma_{wav}(I_i)\right)^2 + \xi^2}, \tag{15}$$

where ξ is a positive parameter, which is set to 10^{-6} in our experiments; Γ_{wav} is the wavelet decomposition function with four decomposed bands.

In the end, we calculate the total loss as follows:

$$Loss_{total} = Loss_{char} + \alpha Loss_{wav} \tag{16}$$

in which $0 < \alpha \leq 1$ is a trade-off factor. In experiments, we set $\alpha = 0.6$ referring to Ref. [3].

3   Experiments

3.1   Datasets and Training Details

We have conducted extensive experiments on two public datasets: the LCD-Moiré dataset [20] and the Sun [19] moiré dataset. Moreover, MSDFE is built by using PyTorch1.9 with CUDA11.0 and accelerated by applying two NVIDIA TITAN_RTX GPUs with data-parallel. Specially, the Adam optimizer is used to minimize the loss function to train the model, and the learning rate with an initial value of 10^{-4} is reduced by 10 times for every 30 epochs. The network is fully trained for 60 epochs.

3.2   Comparison with State-of-the-Arts

To evaluate the gap between generated moiré-free images and ground truths, we use the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as the performance indicators.

We compare our proposed MSDFE with recent deep-learning-based demoiréing methods, including Islab [21], DMCNN [19], MSFE [6], MopNet [7], MDDM [2] and MDDM+ [3]. Besides, we take VDSR [11], DnCNN [22] and UNet [17] into our benchmark test.

Table 1 lists the demoiréing results on the LCDMoiré dataset, where the best and second-best values are in bold and underlined types, respectively. It can be seen on the LCDMoiré dataset, our MSDFE can achieve the best PSNR and SSIM among compared methods. The best PSNR is 44.10 dB obtained by MSDFE, which is 0.66 dB higher than the second-best value. The SSIM of MSDFE is slightly higher than that of MDDM+. Table 2 shows the demoiréing results on the Sun moiré dataset. The results on the Sun moiré dataset indicate that our MSDFE can achieve 30.14 dB in PSNR and 0.897 in SSIM that are significantly better than and compared to the state-of-the-arts, respectively.

Table 1. Comparison of PSNR and SSIM among DnCNN [22], VDSR [11], UNet [17], MSFE [6], DMCNN [19], MDDM [2], Islab [21], MDDM+ [3] and MSDFE on LCDMoiré dataset.

	DnCNN	VDSR	UNET	MSFE	DMCNN	MDDM	Islab	MDDM+	MSDFE
PSNR (dB)	29.08	32.36	34.84	36.66	37.41	42.49	42.90	<u>43.44</u>	**44.10**
SSIM	0.906	0.964	0.971	0.981	0.982	0.994	0.995	<u>0.996</u>	**0.997**

3.3   Visual Results

The visual results of our proposed MSDFE are compared with the recently proposed methods, including DnCNN [22], MSFE [6], DMCNN [19], and MDDM+ [3]. The visual results on some images from the LCDMoiré and Sun datasets are shown in Fig. 5 and Fig. 6, respectively.

Table 2. Comparison of PSNR and SSIM among DnCNN [22], VDSR [11], UNet [17], MSFE [6], DMCNN [19], MopNet [7], MDDM [2], MDDM+ [3] and MSDFE on Sun dataset.

	DnCNN	VDSR	UNET	MSFE	DMCNN	MopNet	MDDM	MDDM+	MSDFE
PSNR (dB)	24.14	24.68	26.49	25.31	26.77	27.75	28.11	<u>30.03</u>	**30.14**
SSIM	0.834	0.837	0.864	0.837	0.871	0.895	0.895	**0.897**	**0.897**

Each row in Fig. 5 shows the moire image, demoiréing results obtained by five methods, and the ground truth from left to right, where the area in a red square is taken as the zoom-in part such that we can compare the details of results. It can be easily found that our proposed MSDFE can restore moiré details and provide the clearest results, and the compared methods have a bad performance for generating blocking artifacts and incorrect color tunes. Furthermore, these methods are ineffective in removing low-frequency moiré patterns.

In Fig. 6, the three images from the Sun dataset are photos taken from real-world objects. Thus, the demoiréing task on the Sun dataset is more difficult than that on the LCDMoiré dataset consisting of computer-rendered images.

We can easily find that our MSDFE recovers correct color tunes, while other methods fail to do although they remove most of the moiré patterns. The above visual results on LCDMoiré and Sun datasets indicate that our proposed MSDFE outperforms the state-of-the-art methods.
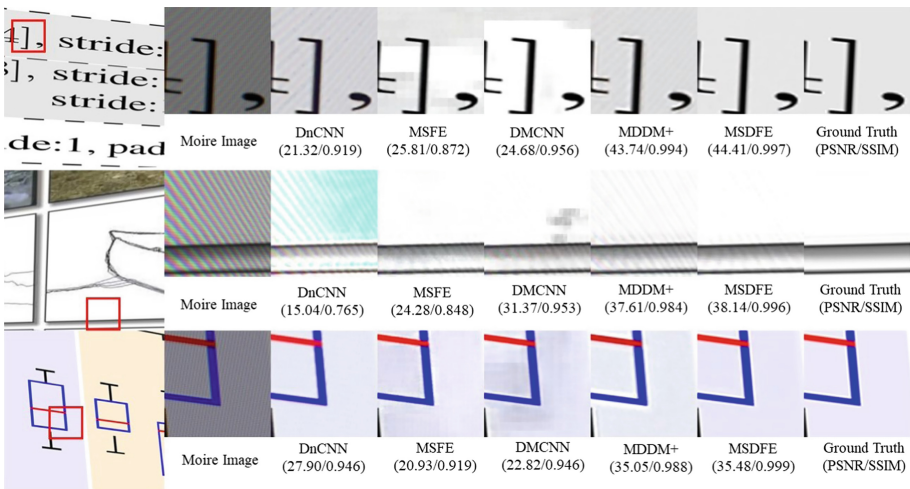
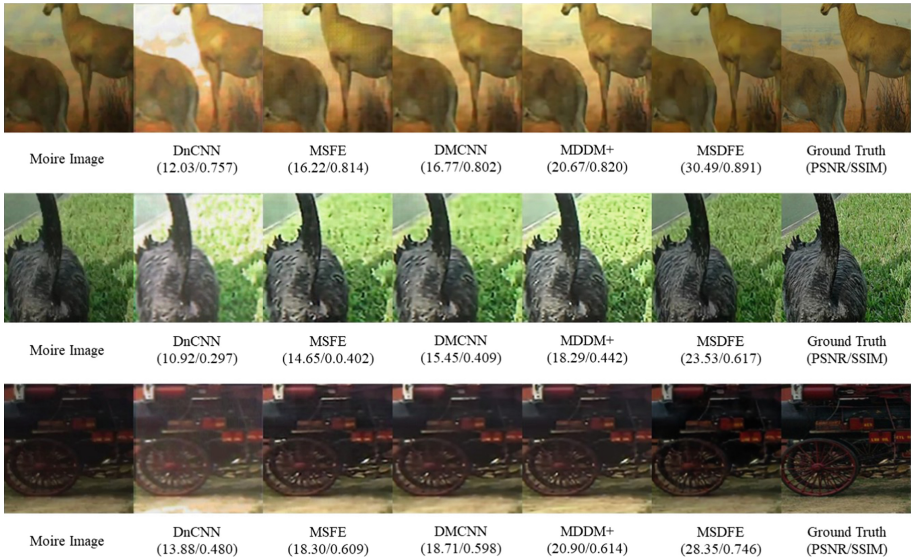**Fig. 5.** Visual comparison on the LCDMoiré dataset.

Fig. 6. Visual comparison on the Sun dataset.

3.4 Model Parameters

In this part, we compare our MSDFE with seven demoiréing methods on the model performance and the model parameters. Parameters are an indicator to measure the size of a model, an efficient network would have an impressive performance with a relatively small number of parameters. Figure 7 shows the performance vs. the number of parameters of compared methods on the LCD-Moiré dataset. From Fig. 7, we can see that MSDFE (red dot) can achieve the best results only with a relatively small number of parameters.
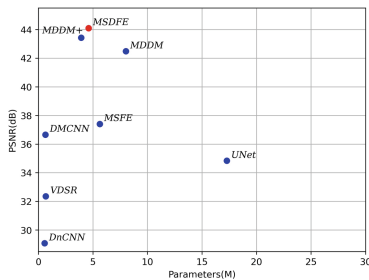

Fig. 7. PSNR vs. parameter number for eight methods on LCDMoiré dataset (Color figure online)
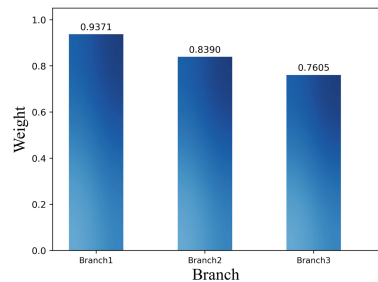

Fig. 8. Weight of branches.

4   Ablation Study

4.1   Network Branches

First, we study the impact of network branches on the performance of our network. In fact, different branches encode image representations at different frequencies and address image features at different resolutions; thus, the branch number in the network greatly influences its performance and size. Table 3 lists the network performance and size on the LCDMoiré dataset, where the size of network is measured by parameter number.

From Table 3, we can know that the network performance gradually improves as the number of branches grows until it reaches 3. After that, the network performance starts decreasing because too many branches cause the loss of detailed information in images. As a result, the network with 3 branches is the optimal choice.

In addition, the weight of each branch can be automatically learned by the scaling module in our network, which can be seen as the indicator of the importance of branches for the final reconstruction of demoiréing image. The weight of each branch is shown in Fig. 8. We can find that Branch 1 is the most important; thus, a branch with higher resolution is of greater importance.

Table 3. Network performance and parameters of MSDFE with different branches.

Branch	PSNR (dB)	Parameter (M)
1	36.04	1.14
2	38.96	2.45
3	44.10	3.91
4	43.71	5.51
5	43.55	7.27

4.2   Multi-spectral Dynamic Feature Encoding

Next, we study the network performance and complexity with and without our proposed multi-spectral dynamic feature encoding. This encoding way is leveraged to encode the dynamic properties of moiré. Experimental results are shown in Table 4. It can be easily found that the PSNR score of the network output is improved from 40.93 dB to 44.10 dB on the LCDMoiré dataset after adding the encoding module. The network parameters also can be found in Table 4. The results indicate that our proposed MSDFE is a lightweight module. MSDFE has a great performance improvement but a small parameter number increase, say 0.55 M.

Table 4. Network performance and parameters of baseline and MSDFE.

Structure	PSNR (dB)	Parameter (M)
Baseline	40.93	3.36
MSDFE	44.10	3.91

5   Conclusion

Demoiréing is more challenging than other image restoration tasks because of its variability and dynamics. In this paper, we propose MSDFE for image demoiréing. The multi-scale network structure of MSDFE can help our model remove moiré in different frequency bands and retain more image details. Besides, we leverage a multi-spectral dynamic feature encoding module to encode moiré patterns dynamically, which makes the model adapt to the transient nature of the moiré patterns. Experimental results indicate that the model achieves 3.17 dB higher in PSNR after adding the multi-spectral dynamic feature encoding module. In addition, our proposed MSDFE significantly outperforms state-of-the-art methods and can remove moiré patterns effectively.

References

1. Ahmed, N., Natarajan, T., Rao, K.R.: Discrete cosine transform. IEEE Trans. Comput. **100**(1), 90–93 (1974)
2. Cheng, X., Fu, Z., Yang, J.: Multi-scale dynamic feature encoding network for image demoiréing. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 3486–3493. IEEE (2019)
3. Cheng, X., Fu, Z., Yang, J.: Improved multi-scale dynamic feature encoding network for image demoiréing. Pattern Recogn. **116**, 107970 (2021)
4. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 11065–11074 (2019)
5. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2015)
6. Gao, T., Guo, Y., Zheng, X., Wang, Q., Luo, X.: Moiré pattern removal with multi-scale feature enhancing network. In: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), pp. 240–245. IEEE (2019)
7. He, B., Wang, C., Shi, B., Duan, L.Y.: Mop moire patterns using mopnet. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 2424–2432 (2019)
8. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)
9. 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, pp. 770–778 (2016)

10. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1501–1510 (2017)
11. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)
12. Kupyn, O., Budzan, V., Mykhailych, M., Mishkin, D., Matas, J.: Deblurgan: blind motion deblurring using conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8183–8192 (2018)
13. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 624–632 (2017)
14. Li, T., et al.: A new design in iterative image deblurring for improved robustness and performance. Pattern Recogn. **90**, 134–146 (2019)
15. Qian, Z., Huang, K., Wang, Q.F., Xiao, J., Zhang, R.: Generative adversarial classifier for handwriting characters super-resolution. Pattern Recogn. **107**, 107453 (2020)
16. Qin, Z., Zhang, P., Wu, F., Li, X.: FcaNet: frequency channel attention networks. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 783–792 (2021)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)
19. Sun, Y., Yu, Y., Wang, W.: Moiré photo restoration using multiresolution convolutional neural networks. IEEE Trans. Image Process. **27**(8), 4160–4172 (2018)
20. Yuan, S., Timofte, R., Slabaugh, G., Leonardis, A.: Aim 2019 challenge on image demoireing: dataset and study. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 3526–3533. IEEE (2019)
21. Yuan, S., et al.: Aim 2019 challenge on image demoireing: methods and results. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), pp. 3534–3545. IEEE (2019)
22. 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. **26**(7), 3142–3155 (2017)