Multiscale Coarse-to-Fine Guided Screenshot Demoiréing

Duong Hai Nguyen, *Graduate Student Member, IEEE*, Se-Ho Lee, *Member, IEEE*, and Chul Lee, *Member, IEEE*

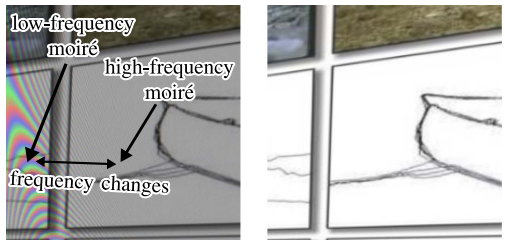
***Abstract*—In this letter, we propose a multiscale coarse-to-fine guided screenshot demoiréing algorithm. We first extract the multiscale features of the input image. Then, we develop the multiscale guided restoration block (MGRB), which removes moiré patterns with the guidance of multiscale information by exploiting the correlation between moiré frequencies. To this end, we design two blocks for feature modulation and moiré pattern removal. In addition, to further improve the performance, we develop an adaptive reconstruction loss to direct the network to focus on regions that are difficult to restore. Experimental results on multiple datasets demonstrate that the proposed algorithm provides comparable or even better demoiréing performance than state-of-the-art algorithms.**

***Index Terms*—Image demoiréing, convolutional neural networks (CNNs), image restoration.**

I. Introduction

When taking pictures of screens with a digital camera, frequency aliasing between the camera sensors and the screen's subpixels produces colorful artifacts in the captured images, called moiré artifacts [1]. As shown in Fig. 1(a), moiré artifacts appear as stripe-shaped color patterns across the image that severely deteriorate the perceptual quality of the images. Furthermore, moiré artifacts may degrade the performance of subsequent computer vision tasks. Therefore, research on screenshot demoiréing, which aims to recover underlying clean images from degraded moiré images, has been recently conducted to enhance the image quality of screenshots.

Recently, deep learning-based approaches that employ convolutional neural networks (CNNs) [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14] have been actively developed and achieved significant demoiréing performance. For example, Sun et al. [1] proposed one of the first multiscale CNN-based screenshot demoiréing algorithms for coarse-to-fine feature handling, demonstrating the effectiveness of multiscale processing in screenshot demoiréing. Because of its effectiveness, various multiscale network-based demoiréing algorithms have been sebsequently proposed [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. In addition, He et al. [13] developed a multimodal learning-based approach, which exploits multiple attributes, including edge information, moiré frequency, shape, and colors, to better model moiré patterns. To improve demoiréing performance by addressing the irreversible issue of convolution, the frequency domain information as well as spatial information has been exploited in [8], [9], [10], [11], [15]. In particular, Zheng et al. [8], [9] developed a demoiréing algorithm using a learnable bandpass filter to exploit the various moiré artifacts in the frequency domain. However, since CNN-based algorithms are based on supervised learning, they require a large amount of clean and moiré image pairs as training data. To address this limitation, unsupervised learning-based approaches that use generative adversarial networks (GANs) with unpaired datasets have recently been developed [16], [17], [18]. They estimate the clean and moiré distributions through adversarial learning to generate demoiréd images via image-to-image translation or to construct a pseudo-paired dataset for supervised training.



(a) Moiré image (b) Clean image

Fig. 1. Example of a moiré image and the corresponding clean image.

Despite significant improvements in the demoiréing performance, recent approaches do not yet address two key moiré properties, which are illustrated in Fig. 1(a). First, moiré patterns are unevenly distributed over an image and vary in shape, color, and frequency [1]. Conventional approaches, however, treat regions with different levels of moiré artifacts equally, which may result in performance degradation. Second, moiré patterns with different frequencies exhibit a strong correlation, as their frequencies gradually change in adjacent areas. Although the information pertaining a particular range of frequencies provides information on other frequencies, conventional algorithms combine multiscale features through simple addition or concatenation; thus, this approach may fail to accurately capture their correlations.

In this work, we propose a multiscale coarse-to-fine guided screenshot demoiréing algorithm to address the limitations of conventional algorithms and better exploit moiré properties.

Manuscript received 1 April 2023; revised 14 June 2023; accepted 10 July 2023. Date of publication 17 July 2023; date of current version 26 July 2023. This work was supported by the National Research Foundation of Korea (NRF) funded by the Korea Government (MSIT) under Grants 2022R1F1A1074402 and RS-2022-00165700. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Kunal Narayan Chaudhury. *(Corresponding author: Chul Lee.)*

Duong Hai Nguyen and Chul Lee are with the Department of Multimedia Engineering, Dongguk University, Seoul 04620, South Korea (e-mail: duongnguyen@mme.dongguk.edu; chullee@dongguk.edu).

Se-Ho Lee is with the Department of Information and Engineering, Jeonbuk National University, Jeonju-si 54896, South Korea (e-mail: seholee@jbnu.ac.kr).

Digital Object Identifier 10.1109/LSP.2023.3296039

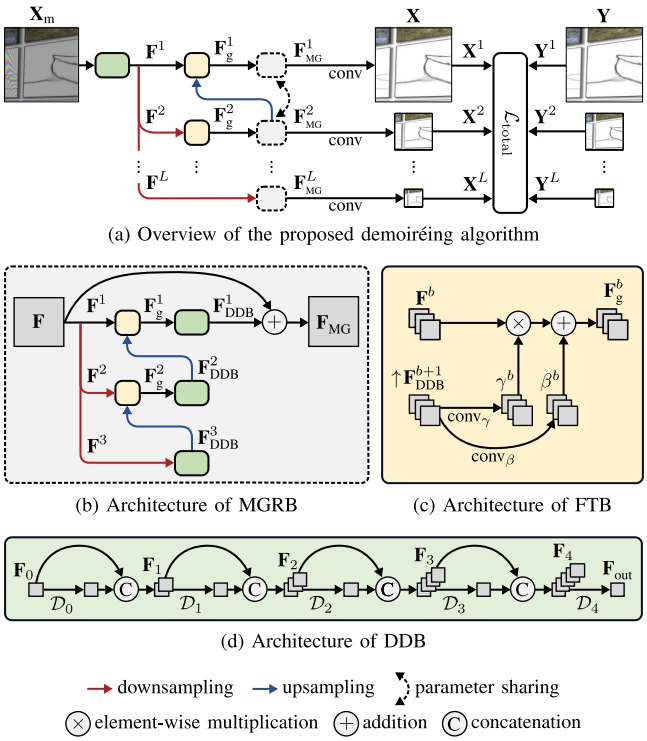


Fig. 2. Overview of the proposed multiscale guided coarse-to-fine network.

The proposed algorithm first extracts the multiscale features of the input image to construct a feature pyramid. Then, we develop multiscale guided restoration blocks (MGRBs) at each level to remove moiré artifacts guided by the restored feature map from the previous level. Each MGRB consists of three branches, each of which handles the moiré patterns in a specific range of frequencies using two blocks: the feature transformation block (FTB), which accounts for the correlation between moiré frequencies by feature modulation, and the dense demoiré block (DDB), which removes moiré patterns in the features. In addition, we develop an adaptive loss that directs the network to focus on regions with denser moiré artifacts. Experimental results demonstrate that the proposed algorithm provides restored images of comparable to or even higher quality than those of state-of-the-art demoiréing algorithms.

In summary, we make the following contributions.

- We propose a multiscale screenshot demoiréing algorithm that learns the correlations between moiré frequencies through multiscale guidance.
- We develop a new adaptive reconstruction loss to focus on denser moiré regions by exploiting the uneven distribution of moiré patterns across the image.
- We experimentally show that the proposed algorithm provides comparable or even better demoiréing performance than state-of-the-art algorithms on multiple datasets.

II. Proposed Algorithm

Fig. 2(a) shows an overview of the proposed algorithm, which consists of MGRBs, FTB, and DDB, detailed in Figs. 2(b), (c), and (d), respectively. Given a moiré image $\mathbf{X}_m \in \mathbb{R}^{H\times W\times 3}$, where H and W are the height and width, respectively, the proposed algorithm aims to obtain a clean image $\mathbf{X}$ similar to the ground-truth $\mathbf{Y}$. The proposed algorithm first extracts feature map $\mathbf{F}^1 \in \mathbb{R}^{H\times W\times C}$, where $C = 144$ is the number of channels, using a DDB. Then, an L-level feature pyramid $\{\mathbf{F}^l\}_{l=1}^L$ is constructed by recursively downsampling $\mathbf{F}^1$ by a factor of two using bilinear interpolation. Next, at each level l, a feature map is restored by suppressing moiré components in the input image spread over a broad spectrum of frequencies. Specifically, except for the coarsest level, each feature map $\mathbf{F}^l$ is guided by the restored feature of the previous level $(l+1)$ using an FTB to obtain $\mathbf{F}_g^l$, which is then fed into an MGRB that learns to suppress moiré artifacts. All MGRBs at different levels share their network parameters to consider the correlations between moiré patterns across multiple scales; this significantly reduces the number of parameters. Finally, the demoiréd image $\mathbf{X}^l$ is obtained using a convolution layer. We present the key components of the proposed algorithm in detail.

A. Multiscale Guided Restoration Block (MGRB)

The MGRB shown in Fig. 2(b) comprises three branches, where each branch $b \in \{1, 2, 3\}$ handles moiré features at a specific scale, and a skip connection. Specifically, given an input feature map $\mathbf{F}_g^l$, similar to the overall architecture in Fig. 2(a), MGRB generates a feature pyramid $\{\mathbf{F}_g^{l,b}\}$ via recursive bilinear interpolation. Henceforth, we omit the level index l for simpler notation. Then, at each branch b, a feature map $\mathbf{F}_g^b$ is concatenated with the output feature of the previous branch and then fed into a DDB that suppresses moiré components to obtain $\mathbf{F}_{\text{DDB}}^b$. The coarsest-scale branch removes large-scale, or equivalently low-frequency, moiré patterns and generates global features [19], whereas the finest-scale branch suppresses high-frequency patterns. Therefore, the MGRB provides guidance for demoiréing across multiple scales by supplying clues from one branch to the next. Finally, at the finest branch, the output feature map $\mathbf{F}_{\text{MG}}$ is obtained by employing a skip connection [20] that combines the input feature $\mathbf{F}_g^1$ and the output of the DDB $\mathbf{F}_{\text{DDB}}^1$, given by

$$\mathbf{F}_{\text{MG}} = \mathbf{F}_g^1 + \mathbf{F}_{\text{DDB}}^1. \tag{1}$$

B. Feature Transformation Block (FTB)

As mentioned previously, moiré frequencies gradually change in adjacent areas from low to high and vice versa with strong correlations between different frequency bands, making moiré patterns sensitive to changes in spatial resolution. Therefore, to exploit the correlation between moiré frequencies during demoiréing, we develop the FTB that modulates features by applying a learnable affine transform [21], which is shown in Fig. 2(c). Specifically, through feature scaling and shifting, high-frequency moiré patterns in finer branches are preprocessed using low-frequency moiré patterns from coarser branches. Thus, the FTB can exploit the correlation between moiré frequencies more effectively than previous approaches, such as concatenation and addition [22], [23]. The FTB is used both in the demoiréing network and MGRB in Figs. 2(a) and (b), respectively. The modulated feature is then fed to the MGRB in Fig. 2(a) and to the DDB in Fig. 2(b). Thus, the FTB uses the demoiréd features from the previous level or scale as guidance information.

As the use of FTB is identical for both the demoiréing network and MGRB, as shown in Figs. 2(a) and (b), respectively, we only describe the FTB for MGRB. Specifically, at each branch

b, we first estimate the transformation parameters $\boldsymbol{\gamma}^b$ and $\boldsymbol{\beta}^b$ using a single convolution layer from upsampled $\mathbf{F}_{\text{DDB}}^{b+1}$, *i. e.*, $\boldsymbol{\gamma}^b = \text{conv}_{\boldsymbol{\gamma}}(\uparrow \mathbf{F}_{\text{DDB}}^{b+1})$ and $\boldsymbol{\beta}^b = \text{conv}_{\boldsymbol{\beta}}(\uparrow \mathbf{F}_{\text{DDB}}^{b+1})$, where $\uparrow$ denotes the upsampling operator using bilinear interpolation. Then, the guided feature $\mathbf{F}_g^b$ is obtained by scaling and shifting the feature map [21] as

$$\mathbf{F}_g^b = \boldsymbol{\gamma}^b \odot \mathbf{F}^b + \boldsymbol{\beta}^b, \tag{2}$$

where $\odot$ denotes the element-wise multiplication.

C. Dense Demoiré Block (DDB)

The DDB is used to extract shallow features in Fig. 2(a) and to suppress moiré artifacts in Fig. 2(b). Fig. 2(d) shows the architecture of the DDB, which is designed based on a dense block [24] that has proved effective in image restoration [25]. In this work, to achieve a superior performance versus efficiency trade-off to those in [26], [27], we concatenate the features of only the current and previous layers. Specifically, it consists of five dilated convolution layers $\mathcal{D}_i$ [28], where $i = 0, \ldots, 4$, each of which is followed by a ReLU activation function. The dilation rates for $\mathcal{D}_i$ are set to 1, 2, 3, 2, and 1, respectively. In addition, unlike the original dense block, which connects each layer to all preceding layers, we only concatenate the features of the current layer with those of the previous layer. More specifically, given an input feature map $\mathbf{F}_0$, the output feature map $\mathbf{F}_{\text{out}}$ is given by

$$\begin{aligned} \mathbf{F}_{i+1} &= \mathcal{D}_i(\mathbf{F}_i) \oplus \mathbf{F}_i, \quad i = 0, \ldots, 3 \\ \mathbf{F}_{\text{out}} &= \mathcal{D}_4(\mathbf{F}_4), \end{aligned} \tag{3}$$

where $\oplus$ denotes the concatenation.

D. Loss Functions

Because moiré patterns are distributed unevenly across an image, as shown in Fig. 1(a), processing each region with the same importance may degrade performance. Thus, inspired by the focal loss [30], we develop an adaptive reconstruction loss that directs the network to focus on difficult-to-recover regions, such as those with dense moiré artifacts. To this end, we first determine the pixels of interest based on their rarity in the input image and the differences between the restored and ground-truth pixel values. Specifically, let $\mathbf{Y}_p$ denote the pixel value at location p in the ground-truth $\mathbf{Y}$ and $h(\mathbf{Y}_p)$ the number of pixels with $\mathbf{Y}_p$. Then, we define the pixel rarity at p as

$$R(p) = 1 - \frac{h(\mathbf{Y}_p)}{|\mathbf{Y}|}, \tag{4}$$

where $|\mathbf{Y}|$ denotes the number of elements in $\mathbf{Y}$. In addition, the difference in pixel values between the demoiréd image and ground-truth also affects the difficulty of restoration. To quantify its effects, we define the pixel value distance at p as

$$D(p) = \lambda^{|\mathbf{X}_p - \mathbf{Y}_p|}, \tag{5}$$

where $\lambda > 1$ is a hyperparameter that controls the sensitivity to the pixel value difference.[1]

As the pixel rarity $R(p)$ and distance $D(p)$ of a region increase, its restoration becomes more challenging. Therefore, we define the adaptive loss $\mathcal{L}_a$ using $R(p)$ and $D(p)$ as weights so that it focuses on difficult regions as

$$\mathcal{L}_a(\mathbf{X}, \mathbf{Y}, \eta) = \frac{1}{|\mathbf{Y}|} \sum_p \left(R(p)D(p)\right)^{\delta(\eta)} |\mathbf{X}_p - \mathbf{Y}_p|, \tag{6}$$

where $\delta(\eta) = \log_{\frac{\eta_{\max}}{\eta_{\min}}}(\frac{\eta_{\max}}{\eta})$ is a function that maps the learning rate η to a value in [0, 1], and $\eta_{\max}$ and $\eta_{\min}$ are the predefined maximum and minimum learning rates, respectively. The learning rate η gradually decreases from $\eta_{\max}$ to $\eta_{\min}$ during training. Therefore, in the first stage, all pixels in $\mathbf{X}$ are treated equally because $\delta(\eta) = 0$. However, as the training process progresses, the network focuses more on difficult pixels since $\delta(\eta)$ approaches to 1.

Finally, we define the total loss $\mathcal{L}_{\text{total}}$ as the sum of the adaptive loss $\mathcal{L}_a$ and perceptual loss $\mathcal{L}_p$ [31] at each level, given by

$$\mathcal{L}_{\text{total}} = \sum_{l=1}^{L} \mathcal{L}_a(\mathbf{X}^l, \mathbf{Y}^l, \eta) + \mathcal{L}_p(\mathbf{X}^l, \mathbf{Y}^l), \tag{7}$$

where the perceptual loss is employed to preserve the high-level similarity between images. More specifically, we compute the ℓ_1 norm between the feature maps of $\mathbf{X}$ and $\mathbf{Y}$ from the seventh layer of the pretrained VGG-16 network [32].

III. Experimental Results

A. Experimental Settings

1) Datasets: We evaluate the proposed algorithm on four public datasets for screenshot demoiréing: LCDMoiré [29], TIP2018 [1], FHDMi [2], and UHDM [3]. The LCDMoiré dataset comprises 10,000 and 100 synthetic moiré and clean image pairs for training and validation, respectively. We use the validation set for the test. The TIP2018 dataset contains 125,000 and 10,000 real image pairs for training and testing, respectively. FHDMi [2] and UHDM [3] contain 9,981 and 2,019 pairs, and 4,500 and 500 pairs of real high-resolution images for training and testing, respectively. We preprocess the data as described in [3].

2) Implementation Details: We train the proposed network with a batch size of 2 using the Adam optimizer [33] on an NVIDIA RTX 3090 GPU. The learning rate η is set to $\eta_{\max} = 2 \times 10^{-4}$ and linearly decayed to $\eta_{\min} = 1 \times 10^{-6}$ with warm restarts [34], as in [3]. Specifically, for LCDMoiré, FHDMi, and UHDM, we reinitialize η every 50 epochs and train the network for 200 epochs. For TIP2018, η is reinitialized every 10 epochs, and the total number of epochs is 80. We set the parameter λ in (5) to 5. Each convolution layer has 144 kernels of size 3×3, except for those that generate $\mathbf{X}^l$. We set the number of levels L to 3 for the low-resolution datasets, LCDMoiré and TIP2018, and 4 for the high-resolution datasets, FHDMi and UHDM. The source code and pretrained models are available on our project website.[2]

B. Comparison With State-of-The-Art Algorithms

Table I quantitatively compares the demoiréing performance using PSNR and SSIM. The proposed algorithm yields the highest PSNR scores on the LCDMoiré and UHDM datasets,

[1] Precisely speaking, it is not a distance, since $D(p) \neq 0$ when $\mathbf{X}_p = \mathbf{Y}_p$.

[2] [Online]. Available: https://nhduong.github.io/guided_demoireing_net

TABLE I
QUANTITATIVE COMPARISON OF DEMOIRÉING PERFORMANCES USING PSNR AND SSIM ON DIFFERENT DATASETS

Dataset	DMCNN [1]	MDDM [6]	WDNet [15]	MopNet [13]	MBCNN [9]	FHDe²Net [2]	ESDNet [3]	ESDNet-L [3]	Proposed
LCDMoiré [29]	35.48 / 0.9785	42.49 / 0.9940	29.66 / 0.9670	- / -	45.08 / **0.9967**	41.40 / -	44.83 / 0.9963	45.34 / 0.9966	**45.73** / **0.9967**
TIP2018 [1]	26.77 / 0.8710	- / -	28.08 / 0.9040	27.75 / 0.8950	**30.41** / 0.9000	27.78 / 0.8960	29.81 / 0.9160	30.11 / **0.9200**	30.13 / **0.9200**
FHDMi [2]	21.54 / 0.7727	20.83 / 0.7343	- / -	22.76 / 0.7958	22.31 / 0.8095	22.93 / 0.7885	24.50 / 0.8351	**24.88** / **0.8440**	24.82* / 0.8426*
UHDM [3]	19.91 / 0.7575	20.09 / 0.7441	20.36 / 0.6497	19.49 / 0.7572	21.41 / 0.7932	20.34 / 0.7496	22.12 / 0.7956	22.42 / 0.7985	**22.48*** / **0.8001***
FLOPs (T)	0.41	0.97	-	6.26	4.22	11.41	0.56	0.92	1.71 / 1.73*
No. Params (M)	1.43	7.64	3.36	58.57	14.19	13.57	5.93	10.62	5.66 / 6.18*

For each metric, the best result is shown in boldface, whereas the second-best is underlined. The * symbol indicates results for $L = 4$.

(a) Input (b) DMCNN (c) MDDM (d) WDNet (e) MopNet (f) MBCNN (g) FHDe²Net (h) ESDNet (i) ESDNet-L (j) Proposed (k) GT

Fig. 3. Qualitative comparisons of demoiréing results on the UHDM dataset. The magnified areas within the red rectangles are shown for details.

and the second-best scores on TIP2018 and FHDMi. In addition, the proposed algorithm outperforms all conventional algorithms in terms of SSIM scores on all datasets, except for FHDMi. As MBCNN is designed for low-resolution datasets, its receptive fields are small; thus, it provides superior performance on the LCDMoiré and TIP2018 datasets, which contain low-resolution images, while leading to poor performance on the high-resolution FHDMi and UHDM datasets. These results indicate that the proposed algorithm achieves the best overall performance by effectively removing moiré artifacts at multiple scales with guided features. Table I also compares the computational complexity in terms of the number of floating-point operations (FLOPs) required to process a 1920×1080 image and the number of trainable parameters.

Fig. 3 compares the demoiréing results obtained by each algorithm for the test images in the UHDM dataset. In Figs. 3(b)–(f), DMCNN, MDDM, WDNet, MopNet, and MBCNN, which do not consider high-resolution moiré, fail to remove severe moiré patterns. In Figs. 3(g)–(i), FHDe^{2}Net, ESDNet, and ESDNet-L provide improved results by handling moiré patterns at different scales. However, their results still contain small moiré patterns, as shown in the magnified regions. By contrast, the proposed algorithm in Fig. 3(j) provides the best demoiréd results completely removing any visible moiré patterns, which confirms the effectiveness of the proposed multiscale feature guidance for demoiréing.

C. Component Analysis

We analyze the effectiveness of the FTB and adaptive loss $\mathcal{L}_a$ on demoiréing performance by training the proposed algorithm with different settings using the LCDMoiré dataset. Table II compares the demoiréing performances. The FTB improves demoiréing performance by processing moiré patterns at different scales with multiscale feature guidance. In addition, the adaptive loss $\mathcal{L}_a$ improves the performance by considering the uneven moiré distribution across an image. Finally, using both the FTB and adaptive loss yields the best performance, indicating their effectiveness.

TABLE II
EFFECTS OF FTB AND $\mathcal{L}_a$ ON THE DEMOIRÉING PERFORMANCE

FTB	$\mathcal{L}_a$	PSNR	SSIM
		45.26	0.9965
✓		45.54	**0.9967**
	✓	45.38	0.9965
✓	✓	**45.73**	**0.9967**

For each metric, the best result is shown in boldface.

IV. CONCLUSION

We proposed a multiscale coarse-to-fine guided screenshot demoiréing algorithm that can learn the correlations between different moiré frequencies. Specifically, we first extracted multiscale features from the input image in a coarse-to-fine manner. Then, we developed the FTB and MGRB, which capture the correlations between moiré frequencies through feature modulation and remove moiré artifacts, at each scale. In addition, we developed an adaptive loss to consider the uneven distribution of moiré patterns. Experimental results on multiple datasets showed that the proposed algorithm provides comparable or even higher demoiréing performance than state-of-the-art algorithms. An important direction for future work is to improve the generalization ability of the proposed algorithm.

REFERENCES

[1] 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.

[2] B. He, C. Wang, B. Shi, and L.-Y. Duan, "FHDe2Net: Full high definition demoireing network," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 713–729.

[3] X. Yu et al., "Towards efficient and scale-robust ultra-high-definition image demoiréing," in *Proc. Eur. Conf. Comput. Vis.*, 2022, pp. 646–662.

[4] S. Yang, Y. Lei, S. Xiong, and W. Wang, "High resolution demoire network," in *Proc. IEEE Int. Conf. Image Process.*, 2020, pp. 888–892.

[5] T. Gao, Y. Guo, X. Zheng, Q. Wang, and X. Luo, "Moiré pattern removal with multi-scale feature enhancing network," in *Proc. IEEE Int. Conf. Multimedia Expo Workshops*, 2019, pp. 240–245.

[6] X. Cheng, Z. Fu, and J. Yang, "Multi-scale dynamic feature encoding network for image demoireing," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 3486–3493.

[7] Y. Guo, C. Ji, X. Zheng, Q. Wang, and X. Luo, "Multi-scale multi-attention network for moiré document image binarization," *Signal Process. Image Commun.*, vol. 90, Jan. 2021, Art. no. 116046.

[8] B. Zheng, S. Yuan, G. Slabaugh, and A. Leonardis, "Image demoireing with learnable bandpass filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 3633–3642.

[9] B. Zheng et al., "Learning frequency domain priors for image demoireing," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 11, pp. 7705–7717, Nov. 2022.

[10] A. G. Vien, H. Park, and C. Lee, "Dual-domain deep convolutional neural networks for image demoireing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1934–1942.

[11] H. Wang, Q. Tian, L. Li, and X. Guo, "Image demoiréing with a dual-domain distilling network," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2021, pp. 1–6.

[12] X. Cheng, Z. Fu, and J. Yang, "Improved multi-scale dynamic feature encoding network for image demoiréing," *Pattern Recognit.*, vol. 116, Aug. 2021, Art. no. 107970.

[13] B. He, C. Wang, B. Shi, and L.-Y. Duan, "Mop moiré patterns using MopNet," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 2424–2432.

[14] H. Yue, Y. Mao, L. Liang, H. Xu, C. Hou, and J. Yang, "Recaptured screen image demoiréing," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 31, no. 1, pp. 49–60, Jan. 2021.

[15] L. Liu et al., "Wavelet-based dual-branch network for image demoiréing," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 86–102.

[16] B. Liu, X. Shu, and X. Wu, "Demoiréing of camera-captured screen images using deep convolutional neural network," Apr. 2018, *arXiv:1804.03809*.

[17] H. Yue, Y. Cheng, F. Liu, and J. Yang, "Unsupervised moiré pattern removal for recaptured screen images," *Neurocomputing*, vol. 456, pp. 352–363, Oct. 2021.

[18] H. Park, A. G. Vien, H. Kim, Y. J. Koh, and C. Lee, "Unpaired screenshot image demoiréing with cyclic moiré learning," *IEEE Access*, vol. 10, pp. 16254–16268, 2022.

[19] J. Wang et al., "Deep high-resolution representation learning for visual recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 10, pp. 3349–3364, Oct. 2021.

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

[21] X. Wang, K. Yu, C. Dong, and C. C. Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 606–615.

[22] Y. Niu, Z. Lin, W. Liu, and W. Guo, "Progressive moire removal and texture complementation for image demoireing," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Jan. 18, 2023, doi: 10.1109/TCSVT.2023.3237810.

[23] K. Jiang et al., "Multi-scale progressive fusion network for single image deraining," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 8346–8355.

[24] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 2261–2269.

[25] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2021.

[26] K. Jiang, Z. Wang, P. Yi, and J. Jiang, "Hierarchical dense recursive network for image super-resolution," *Pattern Recognit.*, vol. 107, Nov. 2020, Art. no. 107475.

[27] R. Li, J. Xie, Y. Lin, T. Tong, W. Zheng, and Q. Gao, "Multi-stage network for single image demoireing," in *Proc. SPIE*, vol. 12158, pp. 62–67, Nov. 2021.

[28] F. Yu and V. Koltun, "Multi-scale context aggregation by dilated convolutions," in *Proc. Int. Conf. Learn. Represent.*, 2016.

[29] S. Yuan et al., "AIM 2019 challenge on image demoireing: Methods and results," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops*, 2019, pp. 3534–3545.

[30] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[31] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 694–711.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[33] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent.*, 2015.

[34] I. Loshchilov and F. Hutter, "SGDR: Stochastic gradient descent with warm restarts," in *Proc. Int. Conf. Learn. Represent.*, 2017.