# Moiré Artifacts Removal in Screen-shot Images via Multiple Domain Learning

An Gia Vien, Hyunkook Park, and Chul Lee
Department of Multimedia Engineering, Dongguk University, Seoul, Korea
E-mail: {viengiaan, hyunkook}@mme.dongguk.edu, chullee@dongguk.edu

***Abstract*—We propose a deep learning-based moiré artifacts removal algorithm for screen-shot images using multiple domain learning. First, we develop the pixel and discrete cosine transform (DCT) networks to estimate clean preliminary images by exploiting complementary information of the moiré artifacts in different domains. Next, we develop a clean edge predictor to estimate a clean edge map for the input moiré image. Then, we propose the refinement network to further improve the quality of the pixel and DCT outputs using the estimated edge map as the guide information and to merge the two refined results to provide the final result. Experimental results on a public dataset show that the proposed algorithm outperforms conventional algorithms in quantitative and qualitative comparison.**

## I. Introduction

Moiré artifacts are referred to as disruptive colorful patterns in images captured by digital cameras. Moiré artifacts appear when frequency aliasing between a camera's color filter array (CFA) and high-frequency scene content occurs [1]. In screen-shot images, the CFA of the camera and the screen's subpixel layout are interposed. The captured images contain moiré artifacts with various shapes and color variations onto images, degrading the quality of the photographs. Thus, many algorithms have been developed to effectively remove moiré artifacts and enhance image quality [1]–[5], called image demoiréing.

One approach is to employ an optical low-pass filter in the camera that sits in front of the image sensor for moiré artifact removal [2]. However, this approach requires special hardware and causes over-smoothing in the results due to the low-pass filtering. Another approach [3] uses multi-scale color gradients to combine the color difference from multiple directions in CFA interpolation. However, their algorithm is based on the assumption that at least one of the color channels contains moiré-free information, which may be violated in practice. In [4], an optimization-based algorithm was developed based on the observation that images with moiré patterns have a sparsity property in the discrete cosine transform (DCT) domain.

Recently, convolutional neural network (CNN)-based approaches [6]–[10] have achieved higher performance than model-based algorithms through learning from large-scale datasets. For example, Sun *et al.* [7] developed a multi-scale convolutional network to remove moiré patterns, while Cheng *et al.* [8] improved on Sun *et al.*'s work [7] by developing a dynamic feature encoding module to deal with the dynamic shapes of moiré patterns by embedding the differences between clean and moiré images. He *et al.* [9] used the edge information and appearance attributes of moiré patterns as additional information. Zheng *et al.* [10] modeled moiré artifacts by learning the frequency prior of the moiré patterns using a learnable bandpass filter. The existing algorithms [7]–[10] remove a different amount of information about moiré artifacts, since they exploit different properties of moiré artifacts in a single domain while each domain can remove complementary information about moiré artifacts.

In this work, we develop a novel demoiréing network that removes moiré artifacts in multiple complementary domains, which consists of the pixel network, DCT network, clean edge predictor, and refinement network. First, we develop pixel and DCT networks to remove moiré artifacts in two complementary domains by processing moiré components in the pixel and frequency domains, respectively. Second, we design the clean edge predictor to estimate edge maps that are then used as guide information to refine the complementary outputs. Third, we develop the refinement network to refine pixel and DCT outputs with estimated clean edge maps. Finally, we obtain an output image by merging the two refined results via the fusion network. Experimental results show that the proposed algorithm provides better demoiréing results than the conventional algorithms [6]–[8], [11].

The remainder of this paper is organized as follows: Section II briefly reviews related work. Section III describes the proposed demoiréing algorithm, and Section IV discusses the experimental results. Finally, Section V concludes the paper.

## II. Related Works

### A. Moiré Artifact Removal

Several algorithms have recently been developed to remove moiré artifacts in captured images. Yang *et al.* [4] proposed a model-based algorithm assuming that moiré patterns are well represented as a sparse matrix in the frequency domain. Recently, deep learning-based algorithms have proved capable of providing higher performance. Sun *et al.* [7] developed a multi-scale network to exploit intrinsic correlations between moiré patterns and image content. Cheng *et al.* [8] further improved the performance by employing adaptive instance normalization based on a dynamic feature encoder.

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea Government (MSIT) (No. NRF-2019R1A2C4069806).

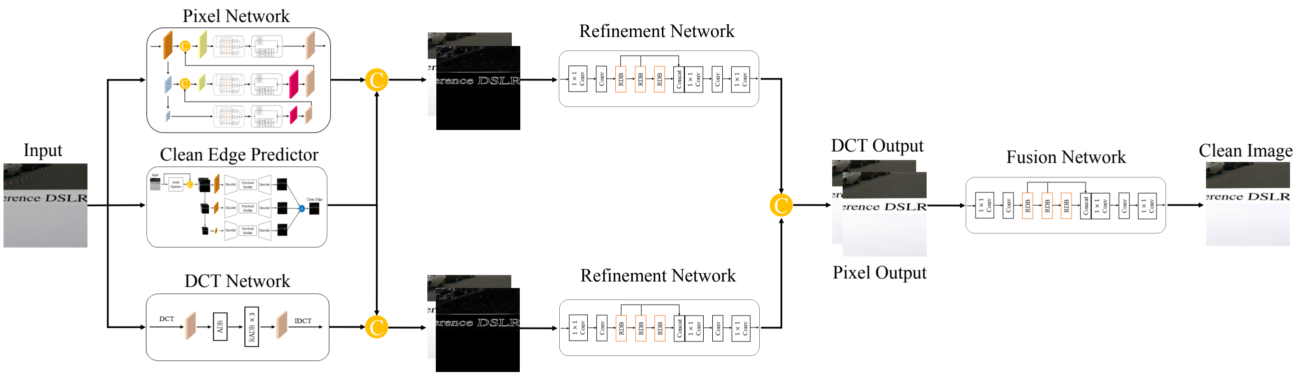



Fig. 1. Overview of the proposed demoiréing algorithm. Two initial demoiréd images are obtained by the pixel network and DCT network, and edge map is estimated by the clean edge predictor. The initial demoiréd images are then refined using the refinement networks. Finally, a pair of refined outputs are fed into the fusion network to yield a clean image.

He *et al.* [9] designed a network that exploits multiple properties of moiré patterns: frequency distribution, edge information, and appearance. Zheng *et al.* [10] developed a network by modeling moiré artifacts in the frequency domain. Note that, conventional approaches have focused on removing moiré artifacts in only a single domain, *i.e.*, the pixel or frequency domain. In this work, we remove moiré artifacts in multiple complementary domains, and then combine their results to obtain better restored images.

### B. Frequency Domain Learning

It has recently been shown that frequency domain processing in CNN-based image restoration approaches provides better reconstruction performance [10]–[13]. For example, Liu *et al.* [12] processed degraded images in the discrete wavelet transform domain for various image restoration tasks, *e.g.*, compression artifacts removal, denoising, and super-resolution. Zheng *et al.* [13] removed compression artifacts in the DCT domain, and then extended the approach to image demoiréing [10]. Vien *et al.* [11] developed a demoiréing network in the DCT domain and combined the output with that of the pixel domain network to improve the demoiréing performance.

## III. Proposed Algorithms

Fig. 1 shows an overview of the proposed network that consists of four main modules: pixel network, DCT network, clean edge predictor, and refinement network. First, we obtain two initial demoiréd results using the pixel and DCT networks. Second, we estimate the clean edge map using the clean edge predictor. Then, each refinement network takes the output of the pixel network or DCT network and uses the estimated edge map as guide information to refine the initial demoiréd results. Finally, a clean image is generated by the refinement network for fusion to a pair of refined results.

### A. Pixel Network

The pixel network processes input moiré images in the pixel value domain. Based on the recent observation that multi-scale contextual information can effectively remove artifacts in images [7]–[10], we adopt a multi-scale approach with three branches corresponding to three different scales in the pixel value domain.

Fig. 2 shows the proposed pixel network with three levels. First, the input image is convolved with a $3 \times 3$ kernel with a stride of 1 to extract the initial features. Then, the feature maps are downsampled from the higher-level to the coarser ones through convolution with a $2 \times 2$ kernel. At each branch, we develop a multiple receptive field block (MRFB) to exploit the properties of moiré patterns better with a larger receptive field and adopt a tone mapping block (TMB) [10] to increase the feature maps' intensities. At the end of each coarse branch, we upsample the features using a single convolutional layer with pixel shuffle [14], and then concatenate them with the input features at the finer scale. Finally, we apply convolution to obtain a final output. The MRFB and TMB are described in detail below.

**Multiple receptive field block:** It has recently been shown that exploring moiré patterns with a larger receptive field is effective for moiré artifact removal [9], [10]. Based on this observation, we develop MRFB with three branches corresponding to three different sizes of receptive field. Fig. 3 shows the structure of the proposed MRFB. To enlarge the receptive field of each branch, we employ dilated convolution [15], where the kernel size is determined by the dilation rate $r$. In this work, we use three different dilation rates $r = \{1, 2, 3\}$ for each branch. In addition, we employ a residual dense block (RDB) [16], [17] to make full use of the hierarchical features extracted from the input image. As shown in Fig. 3, each branch of the MRFB is composed of three parts: feature extraction, RDBs, and features fusion. In particular, we use a single convolutional layer (Conv) to extract the features. These features are fed into the RDBs to extract hierarchical features, and the output features of all RDBs are then fused. Finally, all features from three branches are summed, and a global residual connection is added for stability.

**Tone mapping block:** The moiré image generation model decreases pixel intensities [18], and we assume that the intensity differences can be modeled as a linear transformation.

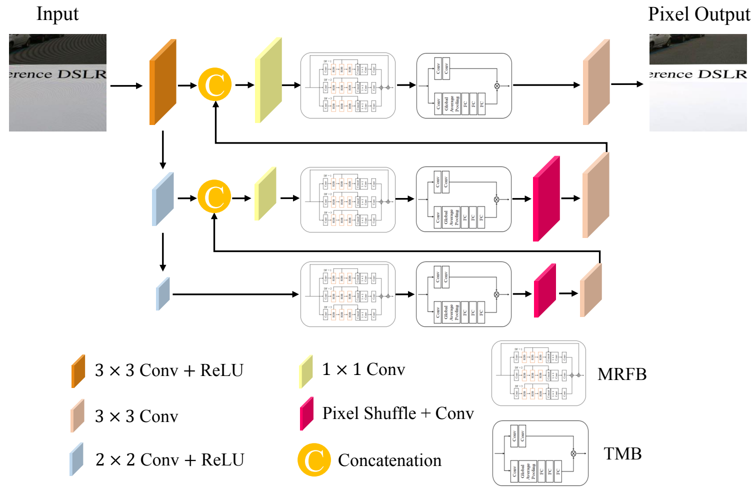



Fig. 2. Overview of the pixel network. The pixel network consists of three branches, each of which removes moiré artifacts associated with specific frequency bands using MRFB and TMB.

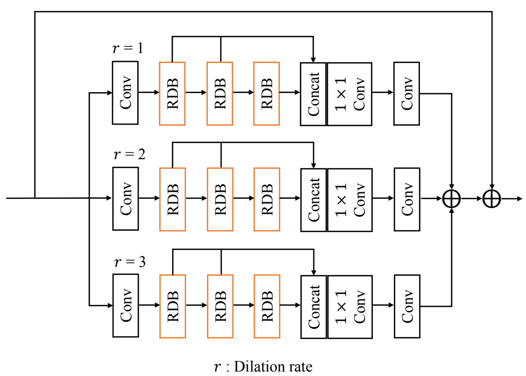



Fig. 3. MRFB consists of three branches with different dilation rates $r = \{1, 2, 3\}$. Each branch is a stack of RDBs with local connections and a global residual connection.

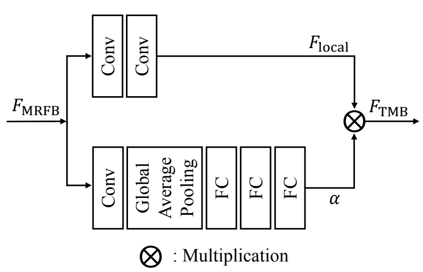



Fig. 4. Structure of TMB to estimate the global parameter $\alpha$.

Thus, to compensate for the intensity reduction by learning global differences, we employ the TMB [10] in this work as shown in Fig. 4. Specifically, the TMB includes global and local branches. In the global branch, given the input feature map $F_{\mathrm{MRFB}}$ obtained from the MRFB, we first extract global features via a $3 \times 3$ convolutional layer with a stride of 1 and a global average pooling. Then, we estimate a global transformation parameter $\alpha$ through two fully connected (FC) layers and another FC layer without an activation function. The local branch extracts the local feature map $F_{\mathrm{local}}$ using two $3 \times 3$ convolutional layers. Then, the output of the TMB $F_{\mathrm{TMB}}$ is obtained by

$$F_{\mathrm{TMB}} = \alpha \times F_{\mathrm{local}}. \tag{1}$$

### B. DCT Network

According to the observation in [9], whereas textures and moiré patterns in an image are hard to distinguish, but the distributions of their transformation coefficients in the frequency domain have different characteristics. Thus, it is necessary to explore the properties of moiré patterns in the frequency domain as a complement to the pixel value domain. Therefore, in this work, we employ the frequency network developed in [11] to process the DCT coefficients.

### C. Clean Edge Predictor

Moiré artifacts have distinct shapes such as curves and stripes. Thus, removing the structure of moiré patterns may erase the texture details in an image. In this work, we first estimate clean edge maps and then use them as guide information to enhance the outputs of the pixel and DCT networks. Fig. 5 shows the proposed clean edge map predictor. First, we extract the initial edge maps using the Sobel operator and concatenate them with the inputs. The clean edge predictor is designed based on U-Net [19] through a multi-scale structure and non-local block [20]. At each branch, the non-local block is employed between the encoder and decoder of the U-Net to capture the semantic edges.

### D. Refinement Network

Since the pixel and DCT networks process moiré images in different domains, they provide results with different reconstructed information that can be used as complementary candidates to produce a final clean image. However, since the results from the pixel and DCT networks retain moiré patterns

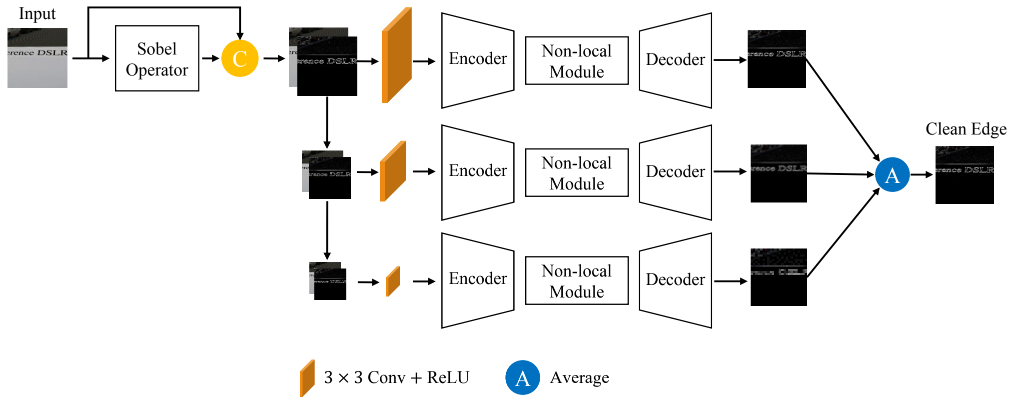



Fig. 5. Overview of the clean edge predictor. Each branch predicts a clean edge map at a specific scale via U-Net [19] and non-local blocks [20].

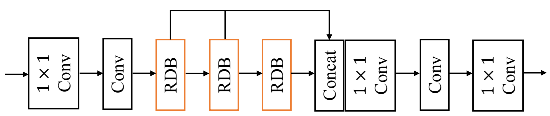



Fig. 6. Architecture of the refinement network.

in highly textured regions, we develop a refinement network to enhance the pixel and DCT outputs with the estimated clean edge map as guide information to better preserve the texture information in the image. Fig. 6 shows the details of the proposed refinement network. Note that the refinement network is also used as the fusion network, as shown in Fig. 1, to obtain a final clean image by merging the two refined results.

### *E. Training*

We have four networks in the proposed algorithm: the pixel network, DCT network, clean edge predictor, and refinement network. We experimentally found that training the networks separately is more effective than end-to-end training in terms of training time and memory usage. Thus, we first train the pixel network, DCT network, and clean edge predictor independently, and then train the refinement networks with the trained networks. We train these networks using the AdamW optimizer [21] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate is initially set to $10^{-3}$ and reduced to one-tenth at every 50 epochs. First, we trained with $128 \times 128$ patches that were randomly cropped from the images, with the batch size of 16. Then, we trained the networks with $256 \times 256$ and then $512 \times 512$ patches for fine-tuning. For the fine-tuning, the learning rate was set to $10^{-5}$, and the batch sizes were set to 8 and 2 for $256 \times 256$ and $512 \times 512$ patches, respectively.

We compute the pixel loss $L_p$ for the pixel network and refinement network as the sum of the $L_1$ loss and the advanced Sobel loss (ASL) [10], given by

$$L_p(\hat{I}, I_{\text{gt}}) = \|\hat{I} - I_{\text{gt}}\|_1 + \lambda \sum_{i=1}^{4} \|S_i(\hat{I}) - S_i(I_{\text{gt}})\|_1, \quad (2)$$

where $\hat{I}$ and $I_{\text{gt}}$ are the output and ground-truth images, respectively, and $\lambda$ denotes a trade-off parameter between the $L_1$ loss and ASL. In addition, $S_i(\cdot)$ denotes the edge map obtained by the $i$th filter in the Sobel filtering among the horizontal, vertical, and two diagonal filters.

For the DCT network, we define the DCT loss $L_{\text{DCT}}$ as the $L_1$ norm between demoiréd and ground-truth images in the DCT domain as

$$L_{\text{DCT}}(\hat{I}, I_{\text{gt}}) = \|T(\hat{I}) - T(I_{\text{gt}})\|_1, \quad (3)$$

where $T(\cdot)$ denotes the DCT operator.

To train the clean edge predictor, we use the $L_1$ norm between the estimated edge map $\widehat{E}$ and the ground-truth $E_{\text{gt}}$, given by

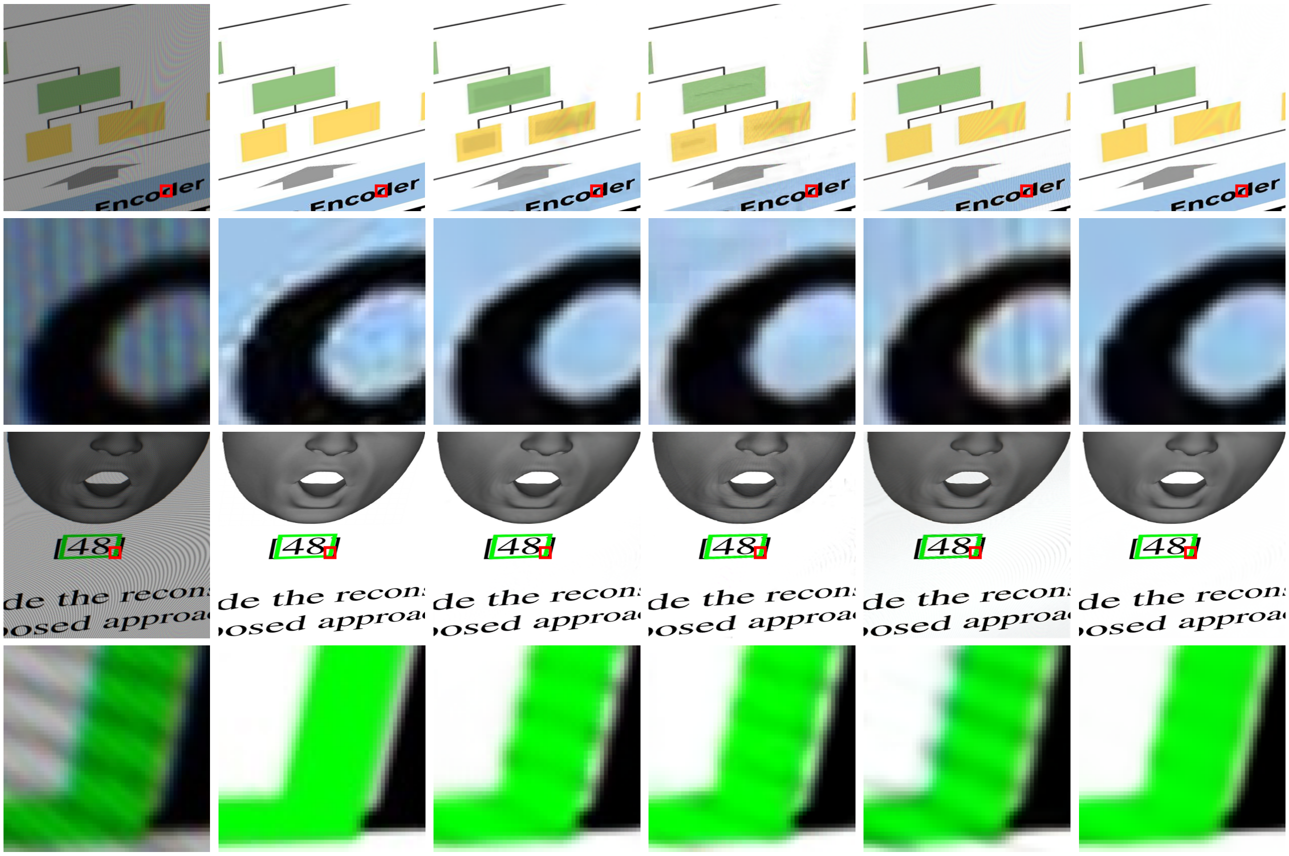
$$L_e(\widehat{E}, E_{\text{gt}}) = \|\widehat{E} - E_{\text{gt}}\|_1. \quad (4)$$

We perform data augmentation to increase the size of the training data. Specifically, we use geometric transformations of $90°$, $-90°$, and $180°$ rotations and horizontal and vertical flipping, thus producing seven additional augmented versions of each image.

## IV. Experimental Results

We evaluate the performance of the proposed algorithm both qualitatively and quantitatively on LCDMoire dataset [5]. The LCDMoire dataset contains 10,200 synthetic moiré and clean image pairs with 10,000 training pairs, 100 validation pairs, and 100 testing images. Since the ground-truth images of the testing set are not released, we use the validation set for the experiments. Note that the validation set is not used in training. We compare the demoiréing performance of the proposed algorithm with those of CAS-CNN [6], DMCNN [7], and DDCNN [11].

Fig. 7 compares the demoiréing results and their detailed parts on the validation set of LCDMoire. The conventional algorithms fail to remove complex moiré patterns. For example, the black stains in Figs. 7(c)–(e) are retained in the green regions. In addition, CAS-CNN and DMCNN in Figs. 7(e) and (d), respectively, cannot faithfully reconstruct the original colors inside the yellow and green rectangles in the top row.



(a) (b) (c) (d) (e) (f)

Fig. 7. Demoiréing results for the validation set (top and third rows) and enlarged regions for the red squares (second and fourth rows): (a) Moiré image, (b) ground-truth, and outputs of (c) CAS-CNN [6], (d) DMCNN [7], (e) DDCNN [11], and (f) the proposed algorithm.

TABLE I
QUANTITATIVE COMPARISON OF CAS-CNN [6], DMCNN [7], DDCNN [11], AND THE PROPOSED ALGORITHM.

| | CAS-CNN [6] | DMCNN [7] | DDCNN [11] | Proposed |
|---|---|---|---|---|
| PSNR | 36.18 | 35.47 | 38.04 | 41.97 |
| SSIM | 0.983 | 0.973 | 0.978 | 0.988 |

In Fig. 7(e), DDCNN provides better color reconstruction but fails to effectively remove moiré artifacts. However, the proposed algorithm outperforms all these conventional algorithms and reconstructs the color information more faithfully. This is because the proposed algorithm benefits from exploiting different properties of moiré images in multiple complementary domains.

Next, in addition to subjective evaluation, we compare the results of the proposed algorithm with those of conventional algorithms objectively using PSNR and the structural similarity index (SSIM) [22]. Table I shows the average PSNR and SSIM results over all images in the validation set of LCDMoire. The proposed algorithm outperforms all the conventional algorithms. Specifically, the proposed algorithm provides a 5.79, 6.50, and 3.93 dB higher PSNR score than DMCNN, CAS-CNN, and DDCNN, respectively. The proposed algorithm also provides the best demoiréing performance in terms of SSIM. This confirms that the proposed algorithm can remove moiré artifacts in images effectively using multiple complementary domain learning and clean edge prediction.

## V. Conclusions

We proposed a deep learning-based moiré artifacts removal algorithm for screen-shot images using multiple domain learning, which consists of the pixel network, DCT network, clean edge predictor, and refinement network. In the pixel and DCT networks, we estimated clean preliminary images by exploiting different characteristics of the moiré artifacts in the pixel and frequency domains, respectively. In the clean edge predictor, we estimated clean edge map for the input moiré image. Finally, the refinement network further improved the quality of the pixel and DCT outputs with an estimated clean edge map as the guide information, and then merged the two refined results to provide the final result. Experiments results on the LCDMoire dataset demonstrated that the proposed algorithm outperforms conventional demoiréing algorithms.

## REFERENCES


[1] S. Yuan *et al.*, "NTIRE 2020 challenge on image demoireing: Methods and results," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2020.

[2] M. Schöberl, W. Schnurrer, A. Oberdörster, S. Fössel, and A. Kaup, "Dimensioning of optical birefringent anti-alias filters for digital cameras," in *Proc. IEEE Int. Conf. Imag. Process.*, Sept. 2010, pp. 4305–4308.

[3] I. Pekkucuksen and Y. Altunbasak, "Multiscale gradients-based color filter array interpolation," *IEEE Trans. Imag. Process.*, vol. 22, no. 1, pp. 157–165, Jan. 2013.

[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. Imag. Process.*, vol. 26, no. 7, pp. 3528–3541, Jul. 2017.

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

[6] L. Cavigelli, P. Hager, and L. Benini, "CAS-CNN: A deep convolutional neural network for image compression artifact suppression," in *Int. Joint Conf. Neural Networks*, May 2017, pp. 752–759.

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

[8] 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*, Oct. 2019, pp. 3486–3493.

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

[10] B. Zheng, S. Yuan, G. Slabaugh, and A. Leonardis, "Image demoireing with learnable bandpass filters," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 3636–3645.

[11] 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*, Jun. 2020, pp. 1934–1942.

[12] P. Liu, H. Zhang, K. Zhang, L. Lin, and W. Zuo, "Multi-level wavelet-CNN for image restoration," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 886–895.

[13] B. Zheng, Y. Chen, X. Tian, F. Zhou, and X. Liu, "Implicit dual-domain convolutional network for robust color image compression artifact reduction," *IEEE Trans. Circuits Syst. Video Technol.*, 2019, accepted.

[14] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1874–1883.

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

[16] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.

[17] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2472–2481.

[18] 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. Workshops*, Oct. 2019.

[19] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Med. Imag. Comput. Computer-Assisted Intervention*, Nov. 2015, pp. 234–241.

[20] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.

[21] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," in *Proc. Int. Conf. Learn. Represent.*, Apr./May 2019.

[22] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Imag. Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.