

# IMAGE DEMOIRÉING WITH A DUAL-DOMAIN DISTILLING NETWORK

Hailing Wang<sup>†</sup>, Qiaoyu Tian<sup>†</sup>, Liang Li<sup>\*</sup>, Xiaojie Guo<sup>\*</sup>

College of Intelligence and Computing, Tianjin University, Tianjin, China  
{whling, joeytian, liangli}@tju.edu.cn, xj.max.guo@gmail.com

## ABSTRACT

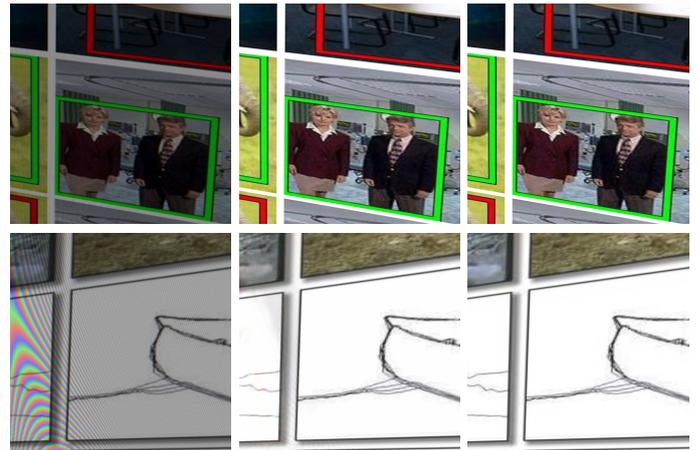
Due to slight discrepancy of spatial frequency between the camera sensor array and sub-pixel layout of LCD monitor, moiré pattern artifacts appear in various shapes and colours which seriously degrade the quality of captured images. It is challenging yet practically crucial to remove moiré artifacts from a single camera-captured screen image. In this paper, we propose a dual-domain distilling network (3DNet for short) to tackle this problem in an end-to-end manner. The 3DNet consists of a dual-branch student network (a.k.a. demoiréing network), and two teacher networks. The two branches of student network exploit knowledge in both spatial-domain and frequency-domain for the sake of removing moiré artifacts, based on the observation that rich image details can be discovered in the frequency-domain while structure information can be well kept in the spatial-domain. The demoiréing process of two branches is supervised by the knowledge distilled from two teacher networks trained for reconstructing clear images in the spatial and frequency domains respectively. Comprehensive experimental results are conducted to demonstrate the efficacy of our design, and reveal its superiority over state-of-the-art alternatives.

**Index Terms**— Moiré Pattern, Demoiréing, Image Restoration

## 1. INTRODUCTION

Recording screen-displayed contents using consumer-level digital devices frequently happens. However, due to interference between sensor array of camera and sub-pixel layout of LCD monitor, moiré artifacts usually show up and severely damage visual quality of captured pictures. These moiré patterns are spatially variant with diverse colour distortion and irregular shapes. Figure 1 (a) shows two such examples contaminated by moiré artifacts. In addition, the moiré pattern spans a wide range of frequency bands, which may cause considerable overlap between moiré artifact and latent image information. Moreover, as no explicit physical models are well-defined for this problem, it is challenging to remove these artifacts by only observing appearance characteristics of moiré pattern.

<sup>†</sup> contributed equally to this work. <sup>\*</sup> corresponding authors.

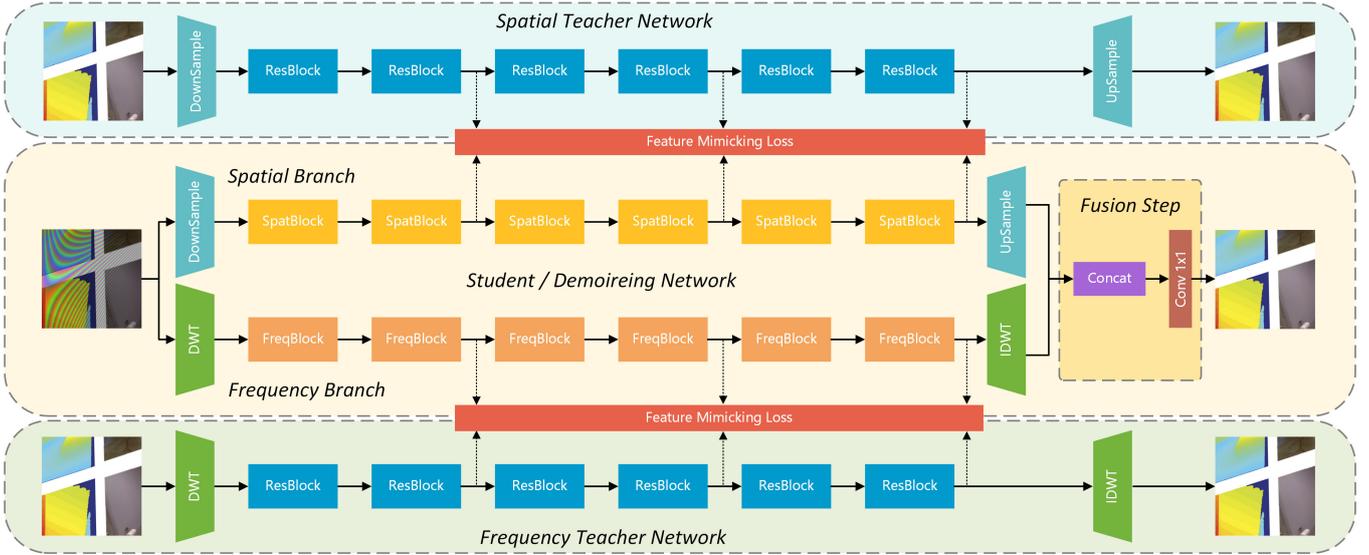


(a) Input image (b) Demoiréing result (c) Ground-truth

**Fig. 1.** Demoiréing results of 3DNet

Recent methods [1][2][3][4][5] treat image demoiréing as an image restoration problem addressed by CNN-based models. For instance, Sun et al. [1] proposed a multi-resolution convolutional neural network, according to the fact that the moiré pattern spans over a wide range of frequency bands, to accomplish the task with a large-scale dataset for demoiréing released. Its follow-ups, with methods [2][5] as representatives, also applied multi-scale based structures to do the job. Besides, Zheng et al. [5] and Liu et al. [4] adopted the frequency prior, which could distinguish moiré patterns from natural image patterns better. He et al. [3] introduced additional labeled data in [1] based on shape, colour, and frequency characteristics for precise moiré pattern removal. Although the above-mentioned methods can produce promising results, knowledge from spatial and frequency domains is barely considered simultaneously. Alternatively, our method works in both the spatial and frequency domains in order to make full use of complementary information from the two domains and obtain better results. In experiments, the effectiveness of this strategy will be verified.

This work introduces a dual-domain distilling network to exploit spatial and frequency prior simultaneously, which is inspired by the success of the previous dehazing method [6].



**Fig. 2.** The overall architecture of proposed framework.

It adopts process-oriented learning mechanism, and builds two networks, say a teacher network for image reconstruction and a student network for dehazing. The similarity of features between teacher and student networks are measured to assist the dehazing task. Similarly, our demoiréing framework also employs the process-oriented learning strategy. We emphasize that, different to their work, our framework includes dual teacher-student pairs specifically designed for two different domains. One teacher is trained for the task of image reconstruction on clear images in spatial-domain, called “Spatial Teacher”. The other is trained for frequency-domain clear image reconstruction named “Frequency Teacher”. The structures of two teachers are largely similar with each other. However, please notice that the “Frequency Teacher” replaces downsampling and upsampling operators by Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT), respectively. Our method employ Haar wavelet transform with level of 2. In terms of the student network, it is constructed as a dual-branch structure to gain knowledge from spatial-domain and frequency-domain at the same time. The learning progress of each branch is guided by the corresponding domain teacher through multi-stage features. It aims at encouraging the student network to learn useful information for demoiréing from the clear intermediate feature representation of the teacher network. Figure 1 illustrates the demoiréing results by our model, from which we can see that the moiré pattern of input can be removed greatly. The main contributions of this paper can be summarized as follows:

- We propose a dual-domain network for demoiréing, which simultaneously considers both spatial-domain and frequency-domain priors for precisely removing moiré patterns from a single image.

- Our model introduces the process-oriented learning strategy to guide the moiré pattern removal process, with a process-oriented loss designed for measuring the similarity of features between teachers and students.
- Extensive experiments are carried out to show the effectiveness of the proposed model compared with other competitors. Besides, ablation studies are also provided to reveal the necessity of each component in our design.

## 2. DUAL-DOMAIN DISTILLING NETWORK

In this section, we will first introduce the dual-domain progress-guided mechanism of teacher-student strategy. After that, the details of our design will be presented.

### 2.1. Dual-Domain Process-Guided Mechanism

The process-oriented learning mechanism was previously applied in dehazing task by Hong et al. [6] with satisfactory performance. The learning process of dehazing network is supervised by the extra network called teacher network in the feature space. In our task, we build two teacher networks to facilitate the training of demoiréing network in two different domains. As shown in Figure 2, two teachers both constitute of a downsampling module, a backbone module, and an upsampling module. The backbone module contains 6 residual blocks [7]. In the frequency teacher network, the downsampling and upsampling modules are replaced by Discrete Wavelet Transform (DWT) and Inverse Discrete Wavelet Transform (IDWT) respectively. Two teacher networks are trained for clear image reconstruction in specific domains so that they can provide student networks with

intermediate feature representation of clear images.

The student/demoiréing network has a similar architecture as the teacher nets except for the backbone module and the fusion step. As shown in Figure 2, the student network consists of two branches, i.e. a spatial branch and a frequency branch. In a few candidate stages of each branch, we encourage their feature maps as close as possible to the clear feature representations of each corresponding teacher network. In order to naturally control the learning process of students, we relax the constraints of the shallow stages while enhancing the deep ones.

## 2.2. Architecture Design

The dual-domain/dual-branch network takes the advantages of two different domains, which can complement each other. In our design, we adopt the structure of two branches for two different domains and capture their processed outputs separately. In the end, we concatenate the two outputs and then fuse them by a simple  $1 \times 1$  convolutional operation as the final demoiréing result of our proposed framework.

### 2.2.1. Spatial Branch

As moiré patterns span a wide range of scales in the spatial domain, we establish each spatial block as a multi-scale architecture which can process moiré patterns at different scales. In this paper, we set the spatial block as a 3-scales structure as shown in Figure 3. The spatial branch of student network is stacked by 6 such spatial blocks with 64 channels, and each feature map produced by downsampling. The top branch of spatial block processes moiré patterns at the original scale and the rest two branches at coarser scales. The two downsampling blocks rescale the original input to half and quarter respectively. After that, the three inputs with different scales are fed into three groups of convolutional layer, followed by the Rectified Linear Unit (ReLU) to capture the output feature maps of each branch. The two coarser outputs pass upsampling layers to make their sizes fit the original size of the top scale. Finally, feature maps from each branch are combined together as the output of the spatial block.

### 2.2.2. Frequency Branch

Recent work [4] found that moiré artifacts are more apparent in certain wavelet subbands, where they can be more easily removed after the wavelet transform. In order to remove the moiré artifacts and restore the details of the background image effectively, we introduce a frequency branch to demoiré network. Each feature map produced by DWT and frequency blocks contains 48 channels. As shown in Figure 4, there are 4 residual blocks with ReLU in previous part of frequency block. Then a channel attention module is applied to the output feature map of the previous part. The attention module employs a global average pooling operation followed by two

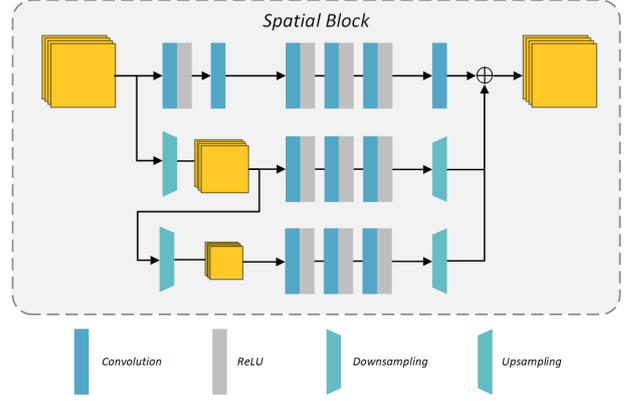


Fig. 3. The spatial block of proposed framework.

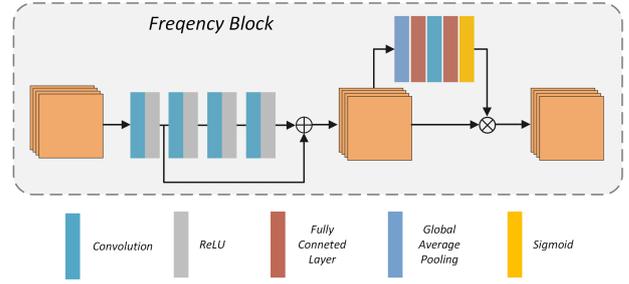


Fig. 4. The frequency block of proposed framework.

fully-connection and a Sigmoid function to learn the weight of each channel. Finally, each channel of the feature map is multiplied by the weight, which means the blocks can automatically select the frequency bands that are most useful for moiré pattern removal through learning.

## 2.3. Objective Function

**Reconstruction loss:** For our framework, the training stage of student network is based on two pre-trained teacher networks. Therefore, in the first stage, the two teacher networks are trained for clear image reconstruction in both spatial-domain and frequency-domain to extract abundant feature representations from clear images. We use L1-distance between the reconstructed result and the ground-truth as reconstruction loss, which is in the following shape:

$$\mathcal{L}_{rec} = \|f_{rec}(I_{gt}; \theta_{rec}) - I_{gt}\|_1, \quad (1)$$

where  $f_{rec}(I_{gt}; \theta_{rec})$  denotes the reconstructed result, and  $I_{gt}$  the corresponding ground-truth image. Once the teacher networks are well-trained, their weights will be frozen. Then, we begin to train the student network.

**Demoiréing loss:** Demoiréing loss measures the distance between the demoiréing result and its ground-truth, which can be formulated as follow:

$$\mathcal{L}_{dem} = \|f_{dem}(I_{moire}; \theta_{dem}) - I_{gt}\|_1, \quad (2)$$

where  $f_{dem}(I_{moiré}; \theta_{dem})$  designates the estimated moiré-free result from the student network.

**Feature Mimicking Loss:** Feature mimicking loss consists of two parts including spatial-domain loss and frequency-domain loss. We introduce the loss to guide the learning process of the student network in the feature space. Each part is defined by a L1-distance between feature maps of each branch of student network and its corresponding domain teacher network in a few candidate stages. To simplify the formula, we let the  $S_m(I_{moiré})$  be the  $m$ -th layer feature map of input with moiré in student network, and the  $T_n(I_{gt})$  the  $n$ -th layer feature map of corresponding clear image in teacher network. The general formula of feature mimicking loss is given in the following:

$$\mathcal{L}_{mim}^{general} = \sum_{(m,n,\mu) \in \mathcal{M}} \mu \|S_m(I_{moiré}) - T_n(I_{gt})\|_1, \quad (3)$$

where  $m, n, \mu$  stand for the  $m$ -layer,  $n$ -layer and weight in the feature mimicking loss of each stage respectively, and  $\mathcal{M}$  is a set of chosen triplets. The total feature mimicking loss includes two parts can be written as follows:

$$\mathcal{L}_{mim} = \mathcal{L}_{mim}^{spatial} + \mathcal{L}_{mim}^{frequency}. \quad (4)$$

**Feature loss:** We adopt the perceptual loss[8] to measure the high-level feature similarity and achieve satisfied performance. The feature loss can be described as the following formula:

$$\mathcal{L}_{feat} = \sum_{l \in \mathcal{N}} \|\Phi_l(f_{dem}(I_{moiré}; \theta_{dem})) - \Phi_l(I_{gt})\|_1, \quad (5)$$

where  $\Phi_l(z)$  is the  $l$ -th layer feature map of  $z$  from pre-trained high-level semantic feature extractor network  $\Phi$ , and  $\mathcal{N}$  denotes a set of chosen layers. In this paper, we adopt VGG-19[9] that trained on ImageNet for image classification as the feature extractor network.

**Overall loss:** During training the student network, the overall objective function is a combination of demoiréing loss in Eq. (2), feature mimicking loss in Eq. (4) and feature loss in Eq. (5). The overall objective function is defined as follows:

$$\mathcal{L}_{overall} = \lambda_1 \mathcal{L}_{dem} + \lambda_2 \mathcal{L}_{mim} + \lambda_3 \mathcal{L}_{feat}, \quad (6)$$

note that the  $\lambda_1, \lambda_2, \lambda_3$  are coefficients that balance the three terms of loss.

### 3. EXPERIMENTS

Our 3DNet is implemented in PyTorch and runs on one Nvidia RTX2080Ti GPU for 36 hours. In our network, the patch size is set to 256 and Adam [10] optimizer with the initial learning rate of  $10^{-4}$  is used. For training of our student network, the learning rate is reduced by half if training loss does not decrease for 5 consecutive epochs.

Model	w/o PGM	w/o $B_S$	w/o $B_F$	w/o CA	Complete
PSNR	36.52	31.81	34.26	34.82	38.05
SSIM	0.985	0.972	0.981	0.982	0.989

**Table 1.** Ablation study of 3DNet for architecture on LCD-Moiré dataset. w/o PGM means training without dual-domain process-guided mechanism. w/o  $B_S$  means training with only frequency branch. w/o  $B_F$  means that only spatial branch is used to train. w/o CA means training without the channel attention modules of frequency blocks.

Model	DnCNN	MSFE	DMCNN	MBCNN	Ours
PSNR	29.08	36.66	36.33	44.04	38.05
SSIM	0.906	0.981	0.980	0.995	0.989

**Table 2.** Quantitative evaluation results compared with DNCNN [12], DMCNN [1], MSFE [2], MBCNN [5].

Model	DMCNN	MBCNN	3DNet
PSNR	17.53	18.18	18.42
SSIM	0.554	0.648	0.579

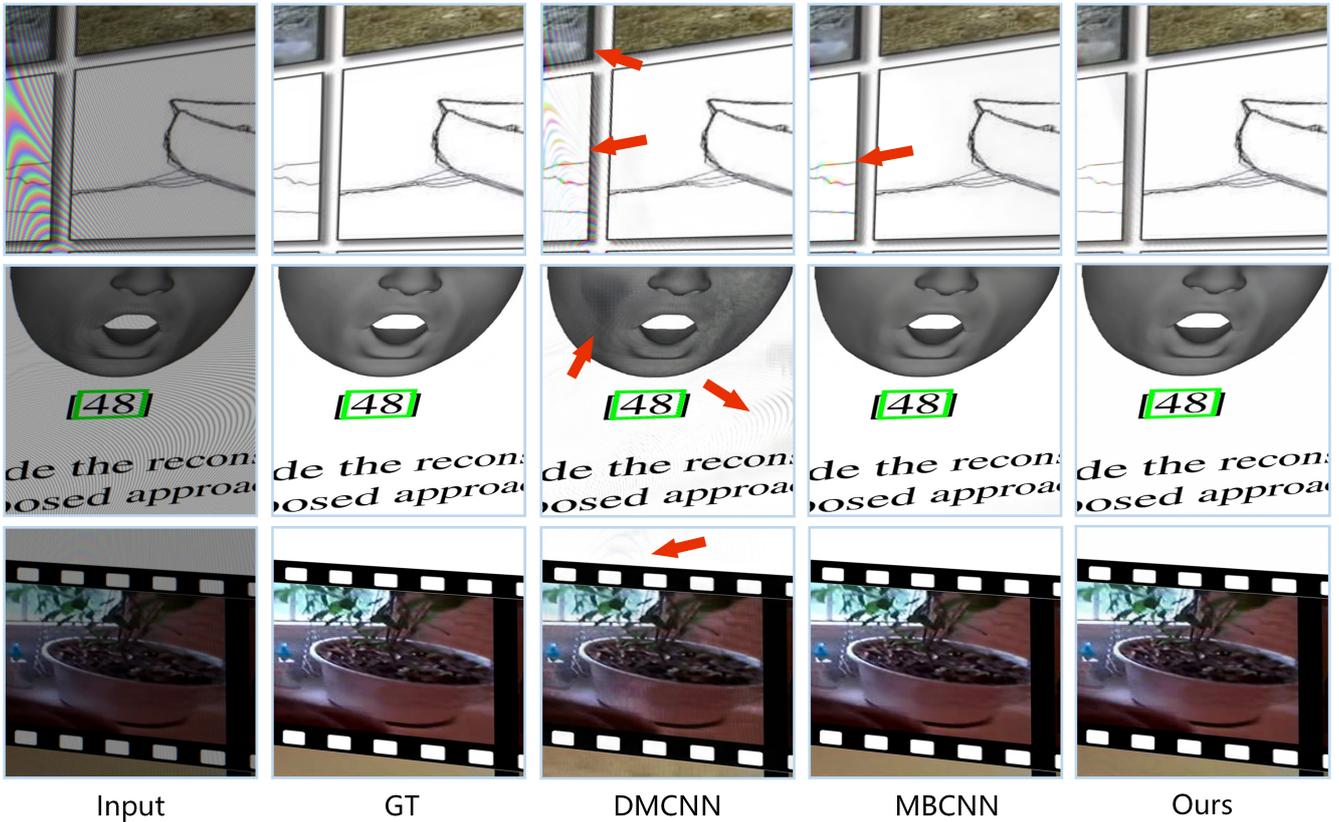
**Table 3.** Quantitative evaluation of generalization ability compared with DMCNN and MBCNN on randomly chosen subset of the TIP2018 dataset.

We compare 3DNet with state-of-the-art methods mainly on the LCDmoiré [11] and TIP2018 [1] datasets and conduct extensive ablation studies to demonstrate the necessity of each component in the design. The LCDmoiré dataset contains 10,200 synthetic image pairs. There are 10,000 training image pairs, 100 validation pairs and 100 testing ones. As the test dataset’s ground truth is not available, the study is conducted on the validation set in this dataset. To verify the generalization ability of our network and others, we randomly select 100 images from the TIP2018 [1] dataset to evaluate. In this dataset, clean images come from the ImageNet ISVRC 2012 dataset and contaminated images are captured by different mobile phones of clean images on different computer screens.

#### 3.1. Ablation Study

To investigate the effectiveness of each component in our model, we conduct the ablation study on teacher network, spatial branch, frequency branch and attention module.

**Teacher network:** This part explores the effect of the proposed teacher networks. The teacher networks guide the training process of student network by a feature mimicking loss which measures the similarity of intermediate features of clear images and demoiréing results. We removed the teacher networks and feature mimicking loss of the 3DNet to disable the guidance of the learning process as a comparison. As shown in Table 1, the teacher network leads to an improve-



**Fig. 5.** Visual comparison of demoiré results among state-of-the-art approaches including DMCNN [1], MBCNN [5] and our method on the samples in LCDMoiré dataset.

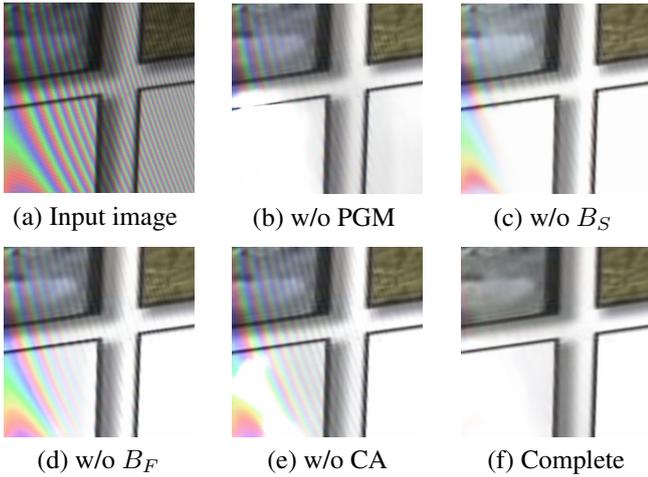
ment of 1.53dB in PSNR. The residual colored stripes shown in Figure 6 indicate the importance of the teacher networks.

**Spatial branch & Frequency branch:** Because the proposed 3DNet is trained with two domains, it is necessary to investigate demoiré results without spatial branch or frequency branch. We cut off the spatial branch and frequency branch respectively to test the ability of the individual branch in 3DNet. Quantitative evaluation results are presented in Table 1. It turns out that the average PSNR value will decrease about 6.24dB without spatial branch and 3.79dB without frequency branch. These results effectively testify the impact of the dual-domain strategy. The results in Figure 6 (c) and (d) visually reflect the importance of the two-domain branches.

**Attention Module:** To see the effect of proposed channel attention module, we built and trained a 3DNet without adding the channel attention module. As shown in Table 1, it turns out that the average PSNR value decreases about 3.23dB. The reason is that the channel attention module receives more evidence to choose which channels should be focused on for moiré pattern removal. Without this channel attention module, our network cannot remove the colored stripe well as can be seen in Figure 6 (e).

### 3.2. Comparison with Prior Work

In this section, we compare the proposed 3DNet with several related models. We choose the widely-used LCDmoiré dataset to quantitatively evaluate different methods. We compare our approach with DnCNN [12], DMCNN [1], MSFE [2], and MBCNN [5]. In Table 2, the quantitative comparisons are reported. From the numbers, we can observe that the DnCNN has limited image demoiré ability compared to other moiré-specific methods. The MSFE and DMCNN proposed to remove moiré in a multi-resolution manner can remove moiré artifacts better than DnCNN. However, these methods ignore the frequency information which is important for moiré removal. Our methods can reach the second highest PSNR, which significantly outperforms DnCNN [12], DMCNN(TIP2018) [1], and MSFE [2], but falls behind the MBCNN. Figure 6 depicts the visual comparison. Our method can remove moiré artifacts in some hard examples such as the first row of Figure 6. Although our 3DNet perform lower PSNR compared with MBCNN [5], the visual difference between our method and MBCNN is hardly viewed in most examples. We notice that our model is of 16.0 Mb, significantly smaller than 54.5 Mb of MBCNN. We also test the generalization ability of our model in comparison with DMCNN [1]



**Fig. 6.** Visual comparison among 3DNet and versions with modifications on LCDmoiré dataset.

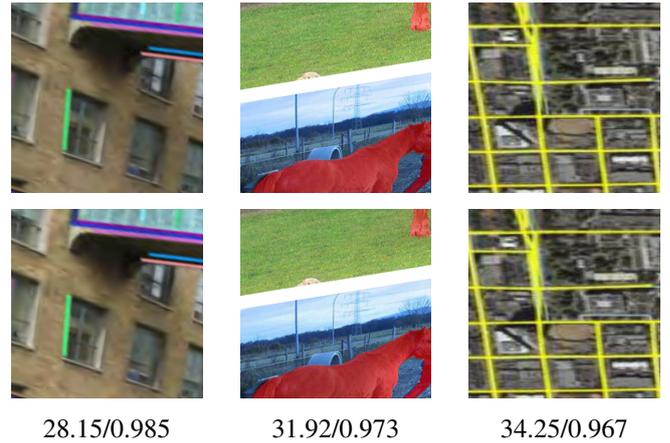
and MBCNN [5]. All the involved competitors are trained on the LCDMoiré dataset but test on the TIP2018 dataset. We randomly choose 100 image pairs of contaminated with moiré patterns and their corresponding clear ground-truth in the TIP2018 dataset. All the above methods are tested on these chosen images and the quantitative evaluation results are shown in Table 3. Although our model achieved the second best SSIM result which is 0.069 lower than the best, our 3DNet beats the MBCNN by 0.24dB in terms of PSNR. Comparing with the result in LCDmoiré dataset, the PSNR of MBCNN [5] is 5.9dB higher than our 3DNet.

#### 4. CONCLUSION

This paper designed a network, namely 3DNet, to remove moiré patterns by considering both the spatial domain and frequency domain priors simultaneously. Moreover, we proposed a dual-domain process-guided learning mechanism for extracting information of clean images from teacher networks. This information could guide the image demoiréing network in the training phase. From extensive experimental results, we can see that our approach achieves overall promising results. In addition, in Figure 7, we present several results with relatively low PSNRs by our method compared with MBCNN. The top row shows the ground truth and the bottom gives our results with PSNR and SSIM, from which we can hardly observe the visual difference between our results and their corresponding ground-truths.

#### Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant nos. 62072327 and 61772512, and TSTC under Grant no. 20JCQNJC01510.



**Fig. 7.** The first row displays three ground-truth images, while the second row gives the recovered results by our method. These results are of relatively low PSNRs but of unaware difference compared with the ground-truths.

#### 5. REFERENCES

- [1] Yujing Sun, Yizhou Yu, and Wenping Wang, “Moiré photo restoration using multiresolution convolutional neural networks,” *IEEE TIP*, 2018.
- [2] Tianyu Gao, Yanqing Guo, Xin Zheng, Qianyu Wang, and Xiangyang Luo, “Moiré pattern removal with multi-scale feature enhancing network,” in *ICMEW*, 2019.
- [3] Bin He, Ce Wang, Boxin Shi, and Ling-Yu Duan, “Mop moire patterns using mopnet,” in *ICCV*, 2019.
- [4] Lin Liu, Jianzhuang Liu, Shanxin Yuan, Gregory Slabaugh, Ales Leonardis, Wengang Zhou, and Qi Tian, “Wavelet-based dual-branch network for image demoiréing,” in *ECCV*, 2020.
- [5] Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Ales Leonardis, “Image demoiréing with learnable bandpass filters,” in *CVPR*, 2020.
- [6] Ming Hong, Yuan Xie, Cuihua Li, and Yanyun Qu, “Distilling image dehazing with heterogeneous task imitation,” in *CVPR*, 2020.
- [7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” in *CVPR*, 2016.
- [8] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, “Perceptual losses for real-time style transfer and super-resolution,” in *ECCV*, 2016.
- [9] Karen Simonyan and Andrew Zisserman, “Very deep convolutional networks for large-scale image recognition,” in *ICLR*, 2015.
- [10] Diederik P Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” *ICLR*, 2014.
- [11] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, and Aleš Leonardis, “Aim 2019 challenge on image demoiréing: Dataset and study,” in *ICCVW*, 2019.
- [12] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, “Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising,” *IEEE TIP*, 2017.