FHDe^{2}Net: Full High Definition Demoireing Network

Bin He[1], Ce Wang[1], Boxin Shi[1,2,3], and Ling-Yu Duan[1,3⋆]

[1] NELVT, Department of CS, Peking University, Beijing, China
[2] Institute for Artificial Intelligence, Peking University, Beijing, China
[3] The Peng Cheng Laboratory, Shenzhen, China
{cs_hebin, wce, shiboxin, lingyu}@pku.edu.cn

Abstract. Frequency aliasing in the digital capture of display screens leads to the moiré pattern, appearing as stripe-shaped distortions in images. Efforts to demoiréing have been made recently in a learning fashion due to the complexity and diversity of the pattern appearance. However, existing methods cannot satisfy the practical demand of demoiréing on camera phone capturing more pixels than a full high definition (FHD) image, which poses additional challenges of wider pattern scale range and fine detail preservation. We propose the Full High Definition Demoiréing Network (FHDe^{2}Net) to solve such problems. The framework consists of a global to local cascaded removal branch to eradicate multi-scale moiré patterns and a frequency based high-resolution content separation branch to retain fine details. We further collect an FHD moiré image dataset as a new benchmark for training and evaluation. Comparison experiments and ablation studies have verified the effectiveness of the proposed framework and each functional module both quantitatively and qualitatively in practical application scenarios.

Keywords: low-level vision, moiré pattern, image restoration.

1 Introduction

The moiré pattern is a widely observed image degradation induced by frequency aliasing between the display and camera. Such an interference between the periodic arrangement patterns of LCD sub-pixels and camera sensors results in conspicuous stripe shaped color distortions across the image, severely deteriorating its visual quality and feature fidelity in visual tasks. Thus demoiréing, indicating the removal of moiré patterns, is an issue of great interest to explore, yet with major challenges.

Demoiréing's challenges reside in the fact that moiré patterns led by camera aliasing can hardly be expressed using an analytical model with a handful of variables. Besides, it is ambiguous to classify the patterns into several typical

⋆ Ling-Yu Duan is the corresponding author.

Fig. 1. 2K resolution (2560×1440) image with moiré pattern degradation (a) and demoiréing result of the proposed FHDe^{2}Net (b). Red and blue boxes provide zoomed-in views of local regions.

categories, considering the diversities in shape and the highly spatially varying structures, as shown in Figure 1. Efforts have been made to alleviate the influence of moiré patterns with both optical filters [19] and post-processing algorithms. Nonetheless, the filters often bring over-smoothing artifacts, thus signal processing based optimization algorithms with the assumptions of sparsity in frequency domain [17] and layered model [33], become more practical solutions. With the surge of deep learning, recent methods [28, 4] exploiting the comprehensive modelling capability of deep features have been proposed, and more promising demoiréing results have been achieved on the data, consisting of images captured from screens [28]. These methods [33, 28, 4] usually take the entire image as direct input for global processing, and work reasonably well on images with limited resolution like 384×384.

However, in the context of the prevailing usage and evolving of camera phones, new challenges have emerged for demoiréing. A prominent issue is the growing resolution of the inputs (*e.g.*, camera phones with FHD, 2K or even higher resolution are mainstream), which cannot be handled by existing methods, especially learning based ones. The challenges brought by the high resolution[4] of images are as follows: 1) High resolution expands the range of the pattern scales, as shown in the comparison between Figure 2 (a) and (b), and commonly used deep networks usually have a total receptive field of about 100×100, which is not sufficiently large for detecting large-scale patterns on full-size high-resolution inputs. 2) An obvious high-resolution detail loss can be observed as a side effect of demoiréing as the example result shown in Figure 2 (c). The over-smoothing distortion is mainly caused by pixel modifications in spatial domain, which ignores the difference between periodic patterns and edges in high-frequency sub-bands. 3) Learning based demoiréing methods suffer performance drop on real high-resolution images. Models trained with existing low-resolution cropped screen images [28] have limited generalization capability on high-resolution images, which is a more practical scenario.

In this paper, we propose the **F**ull **H**igh **De**finition **De**moiréing **Net**work, named **FHDe^{2}Net**, whose framework is shown in Figure 3, to cope with above three challenges in high-resolution image demoiréing: 1) The upper moiré pattern removal branch aims at enlarging the receptive field to address the

[4] Higher than FHD (1920×1080) throughout this paper.

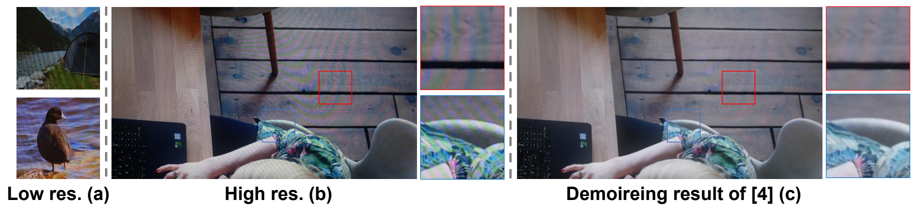

Fig. 2. Comparison between low-resolution image from existing TIP18 benchmark [28] (a), high-resolution moiré image (b), and state-of-the-art [4] result on high-resolution image (c). Moiré pattern residues and over-smoothing distortion can be observed in zoom-in regions marked by red and blue boxes.

expanded pattern scale range in high-resolution images, and eradicating pattern residues, with a cascade of two networks focused on global and local level removal respectively. 2) The lower high-resolution content separation branch is proposed to preserve the fine details against the distortions in processing, by exploiting frequency domain features to disentangle high-frequency contents from moiré patterns. 3) Moreover, we newly collect a Full High Definition Moiré image (FHDMi) dataset as a benchmark to facilitate and evaluate our proposed method. FHDe^{2}Net is verified to deliver satisfying demoiréing results on images with FHD (1080p) or higher resolution, as shown in Figure 1. Our major contributions can be summarized as follows:

- We are the first to explore the emerging problem of high-resolution image demoiréing, and propose a global to local moiré pattern removal strategy to cope with the issues of the wider pattern scale range and demoiréing residues in images of FHD or higher resolution.
- We propose a frequency based high-resolution content separation mechanism, to compensate the fine detail distortions in demoiréing by exploiting signal properties in the frequency domain.
- We contribute the first high-resolution moiré image dataset to benchmark demoiréing tasks, which is composed of FHD camera phone captured screen image pairs. In addition to the high-resolution demoiréing task, we hope our dataset could inspire future research on image restoration towards practical scenarios with latest camera phones.

2 Related work

Moiré pattern removal. The formation of moiré patterns are closely related to the camera imaging process, especially the frequency of the color filter array (CFA). Thus, methods targeted at improving the imaging pipeline have been proposed to eliminate moiré patterns, including anti-aliasing filter on lens [19, 23] and interpolating the output of CFA [18, 21]. However these methods achieve limited success, hence post-processing methods originated from assumed properties of moiré patterns are more frequently adopted for various types of

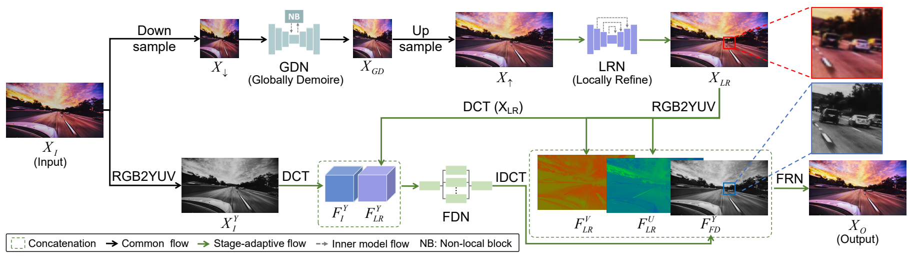

Fig. 3. The framework of proposed FHDe^{2}Net. The input is passed into two branches: The global to local removal branch (upper) is a cascade of Global Demoiréing Network (GDN) and Local Refinement Network (LRN), to eliminate moiré patterns across all scales. High-resolution content separation branch (lower) conserves high-resolution fine details with Frequency based Disentangling Network (FDN). The complimentary outputs of two branches are combined in YUV color space and further refined in Fusion Refining Network (FRN) to deliver the final output.

moiré-contaminated images. Space-variant filters concerning different screening frequencies [27, 25] are proposed for eliminating the simple halftone moiré patterns in scanned images. Liu *et al.* [17, 32] propose a low-rank constrained sparse matrix decomposition method to handle highly textured images. Yang *et al.* [33] propose a layer decomposition model to describe the formation of screen-shot moiré patterns, but at a high computational cost for optimization.

With deep learning booming, demoiréing also starts to benefit from convolutional neural networks recently. Sun *et al.* [28] propose a multi-scale learning strategy with a benchmark dataset captured on LCD screens. He *et al.* [4] improve the learning based methods with property-oriented modules. Generative and adversarial learning framework [8, 15, 34] and synthesized data [6] are also resorted to for removing moiré patterns. These learning based methods have achieved promising results on corresponding testing sets, however they cannot effectively cope with high-resolution images, which are more often confronted in practical applications, especially images captured with latest camera phones.

High-resolution image restoration. Limited research efforts have been dedicated to addressing the high-resolution issue in image restoration. For example. many state-of-the-art image restoration methods for denoising [3, 11], deraining [36], dehazing [12], and reflection removal [30, 39],the targeted inputs are restricted at a relatively low level, ranging from 180×180 to 512×512. The images in corresponding benchmark datasets also have similar limited resolutions [22, 9, 13, 29], with rare exceptions of 720p or higher resolution.

To deal with high-resolution inputs at affordable computational cost, existing computer vision methods often adopt patch based strategy, but has a major drawback of artifacts on patch boundary and low running efficiency (*e.g.*, [5, 31]). Another solution is to downscale the input then conduct super-resolution to the results (*e.g.*, [4]), but such a strategy leads to unavoidable defects including blurry boundaries [35]. A strategy to explicitly deal with high-resolution input for image restoration, particularly the demoiréing task, has yet to be found.

Fig. 4. Visual comparison among images from TIP18 dataset [28] (a), AIM [6] (b) and our proposed FHDMi dataset (c). The multiple curve centers in (c) are denoted by yellow boxes. (d) shows the superimposition result of original region and recaptured FHD image. Please zoom-in to check details.

3 Full high definition moiré image dataset

There exist two datasets serving the task of demoiréing, but neither of them can be applied to benchmarking demoiréing on high-resolution images. The AIM dataset [6], composed of synthesized images based on camera imaging stimulation pipeline, suffers from deviation from real data. The dataset proposed by Sun *et al.* [28] (denoted as TIP18 dataset), contains cropped screen captured real images with limited resolution. This motivates us to create a new **F**ull **H**igh **D**efinition **M**oiré image dataset, named FHDMi dataset.

FHDMi dataset contains 9981 image pairs for training and 2019 for testing. The image pairs are constructed with a moiré-free image as the ground truth, which is the source of moiré image of the same content displayed on screens. The data capture involves various combinations of different models of camera phones and display monitors, for the diversity of data intrinsic distributions[5]. Comparisons among three datasets are shown in Figure 4 and Table 1, and the characteristics of FHDMi dataset are presented as follows:

- **High resolution**: All data in FHDMi dataset have a FHD (1920×1080) resolution, in contrast, the majority of cropped images in existing benchmark TIP18 dataset [28] only have resolution of around 400×400.
- **Pattern complexity in full-screen images**: Moiré patterns in the full-screen captured images of FHDMi dataset contain more diverse and sophisticated structures, like multiple curve centers and streaks of extremely large scale, as shown in Figure 4 (c). Such a complexity can hardly be modelled from cropped images [28] and synthetic data [6].
- **Diverse scenes for practical application**: The ground truth images are collected according to 18 categories of frequently observed contents on screens: wallpapers, sports video frames, film clips, documents, *etc.* In contrast, AIM [6] cannot meet the requirements for real implementation due to the domain gap of homogeneous synthesized data including document screenshots only. TIP18 dataset [28] includes many categories, however does not cover the scenarios concerning screen display like webpages or slides.

Apart from the resolution, another unique characteristic of FHDMi dataset is that we adopt unaligned image pairs. This is because that captured FHD

[5] The detailed settings are presented in the supplementary material

Table 1. Comparison among TIP18 dataset [28], AIM [6] and our proposed FHDMi dataset. "FS" stands for full screen, and "Real" for real captured data.

Dataset	Resolution	Amount	Content	FS	Real
TIP18 [28]	$384 \times 384 \sim 700 \times 700$	135,000	**ImageNet**	×	✓
AIM [6]	1024×1024	10,200	Documents only	×	×
FHDMi	$\mathbf{1920 \times 1080}$	12,000	**Films, sports, etc.**	✓	✓

images tend to contain nonlinear distortions introduced by cameras, which can be visualized in Figure 4 (d), with ghosting edges on the superimposition of two layers. Such distortions inevitably make the accurate pixel-wise calibration and alignment to the original images less reliable. To the best of our knowledge, the FHDMi dataset is currently the only high-resolution benchmark dataset for demoiréing, and it will be publicly available once the paper is published.

4 Methodology

As shown in Figure 3, the framework of FHDe^{2}Net comprises cascaded global to local removal branch, high-resolution content separation branch and a fusion module integrating the intermediate results. The methodology and training details will be introduced in following subsections.

4.1 Cascaded global to local moiré pattern removal

The moire pattern structures vary in terms of scales of the streaks, ranging from thin scanned lines to wide curved stripe regions, and patterns of larger scales are more difficult to eliminate due to their wide coverage and low periodicity. Particularly, in high-resolution images, the scale range is expanded as shown in Figure 2, which makes networks with limited receptive field [4] fail to infer the complete distribution of large-scale patterns, resulting in visible residues after a global-only removal as shown in Figure 2 (c). Patch targeted models breaks the spatial connection across patches, and thus incapable of capturing large-scale patterns across patches, but its focus on local regions can benefit cleaning pattern residues. By taking a trade-off, we propose a cascaded global to local removal strategy to address the pattern scale issues beyond global-only removal and naive patch based methods, as shown in the upper part of Figure 3.

The cascaded branch consists of two sequential parts, the global demoiréing network (GDN) emphasizing on large-scale patterns, and the local refinement network (LRN) to further erase local pattern residues. For GDN, the network takes the downsampled version of the moiré-contaminated high-resolution image X_I as input, denoted as $X_{\downarrow}$, and passes on a dense block based autoencoder with a succession of pooling operations. As such the receptive field of bottleneck neurons of GDN can be consecutively enlarged to more than 400×400 when

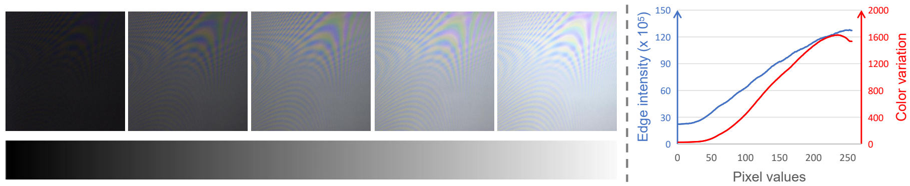

Fig. 5. The observed intensity of moiré patterns, indicated by edge intensity and color variation within pattern regions (left), shows a strong correlation to the brightness of the background (right).

converted back to full high definition size, greatly surpassing common models. Furthermore, to strengthen the internal spatial connection of large-scale moiré patterns on feature maps, non-local blocks [16] are also applied at the bottleneck of GDN, with correlation computation across the feature map. Thereby, with downsampling based receptive field enlarging and non-local features facilitating global removal, the majority of moiré patterns on $X_{\downarrow}$, especially the large-scale ones, can be erased by GDN.

Though GDN works well globally, local pattern residues in its result X_{GD} of GDN still need to be further eliminated by region-targeted LRN. Inspired by the local enhancement strategy in super-resolution methods [10], we adopt a stage-adaptive strategy to make LRN focus on regional refinement, and employ a full convolutional network for the backbone of LRN. The stage-adaptive data flow, denoted by green arrows next to LRN in Figure 3, consists of regions from bilinearly upsampled $X_{\uparrow}$ in training stage, for concentration on learning local residue distributions. In testing stage, the data flow is substituted by the entire $X_{\uparrow}$ for efficient refinement across the image.

We observe that the intensity of moiré pattern, indicated by edge intensity and color variance of moiré covered region, is generally in accordance with the brightness of the region it occupies as shown in Figure 5. Thus to accelerate the learning of local residues for better convergence, we distill the training regions with a mask based selection algorithm. The mask originates from threshold on region brightness using [20], narrowing down potential moiré-sensitive regions, and the masked regions in $X_{\downarrow}$ are selected according to its edge difference to corresponding clean regions, which implies the intensity of moiré residues.

With GDN and LRN cascaded, the output of proposed global to local moiré pattern removal is a pseudo high-resolution moiré-free image X_{LR}, and the overall process can be formulated as follows:

$$X_{LR} = LRN(\uparrow(GDN(\downarrow(X_I)))), \tag{1}$$

where $\uparrow$ and $\downarrow$ denote the upsampling and downsampling operation.

The proposed cascaded global to local moiré pattern removal strategy consequently tackles the challenge of expanded pattern scale range in high-resolution inputs with receptive field enlarging in GDN, and further emphasizes on more delicate residue elimination within local regions by LRN. Apart from erasing residues in a cleaner manner, computational overload caused by high-

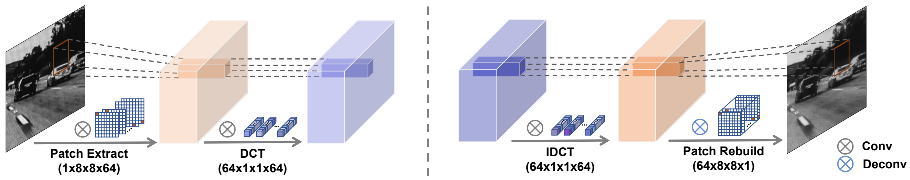

Fig. 6. Illustration of spatial-to-frequency domain transformation (left) and its reverse counterpart (right) realized by convolutions. Each transformation consists of a patch manipulation (patch extraction or rebuilding) and a discrete transformation (DCT or IDCT). Convolution kernels are expressed as $C_{in} \times W \times H \times C_{out}$.

resolution input can also be averted with the separately training GDN and LRN, with downsampled image and cropped regions respectively, which makes the branch potentially capable of handling higher resolution.

4.2 Frequency based high-resolution content separation

Higher resolution gives images the capability to contain more details such as subtle edges and textures. These extra fine details are sensitive to image modifications, thus moiré pattern removal without explicit consideration on content conservation tends to degenerate such high-frequency signals in the image. Moreover, as shown in the red box in Figure 3, high-resolution details are severely lost in X_{LR} due to previous downsampling operation. Therefore, it is necessary to separate the contents with high-resolution details from the original input, to compensate the degradation caused by detail loss. However, how to disentangle high-frequency content from moiré patterns turns out to be the prominent problem. Among moiré patterns, high-frequency patterns within a local region are generally periodical thin lines, which are closely arranged in a unified direction. Such patterns can be easily differentiated from the edges in natural images, considering their periodicity in the frequency domain, since the latter ones are more sparse and diverse in directions.

Therefore, we propose a Frequency based Disentangling Network (FDN) to extract a moiré-free content layer with undistorted high-resolution details, exploiting frequency domain features. We first extract the luminance (Y) channel X_I^Y from the original high-resolution image represented in the YUV color space, since the luminance measures the intensity of light at each pixel according to a particular weighted combination of frequencies. The spatial to frequency domain transformation can be realized by convolution operations [2]. As shown in the left part of Figure 6, the 8×8 overlapped patches in Y channel are first collapsed into 64-dimension vectors with 64 one-hot filters, and then convoluted with $64 \times 1 \times 1$ filters initialized by DCT matrix to complete domain transformation. Hence we can obtain the DCT coefficients of the image across channels as shown in Figure 7, which correspond to different frequency bands. The sub-band coefficients are arranged according to the relative location of patches in obtained feature F_I^Y, thus the spatial relations in images are retained in feature maps, which reasonalizes the subsequent convolutional operations upon F_I^Y.

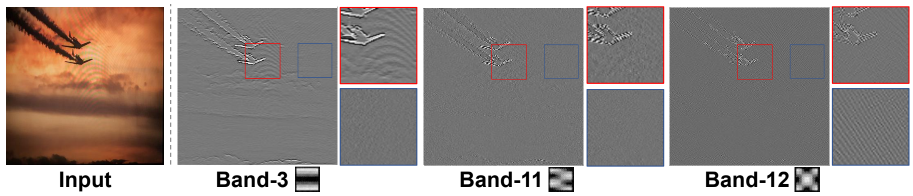

Fig. 7. Visualization of different feature channels after convolutional DCT transformation. Red and blue boxes show zoom-in local regions, where band-3 and band-12 contain patterns with different scales, and band-11 mainly contains high-resolution content details.

Considering the correspondence between feature channels and frequency sub-bands, we adopt the squeeze-and-excitation (SE) block [14, 4] in FDN to learn different weights for each channel to emphasize the disentanglement of high-frequency moiré patterns and image details. Furthermore, to alleviate the difficulty of the disentanglement for patterns of lower frequencies, we introduce guidance from the result of removal branch X_{LR} , which suppresses moiré patterns but lacks high-resolution details. Similarly, we convert the luminance of X_{LR} into DCT representation, then concatenate the guidance F^Y_{LR} with the frequency domain features F^Y_I and integrate them with 1×1 convolution as shown in Figure 3. The integrated features pass through SE blocks and convolutional layers with different dilation sizes, and the multi-scale frequency domain features are further merged and transformed back to targeted content layer with convolutional inverse DCT and patch rebuilding as shown in Figure 6. The obtained moiré-free high-resolution content layer in luminance can be further fused with the color information from pseudo high-resolution result X_{LR}. The overall process of the high-resolution content separation can be presented with the following equation:

$$X^Y_{FD} = \mathcal{D}^*(FDN(\mathcal{D}(X^Y_{LR}) \oplus \mathcal{D}(X^Y_I))), \quad (2)$$

where X^Y_{LR} and X^Y_I stand for luminance channel for corresponding image, $\mathcal{D}$ and $\mathcal{D}^*$ for DCT and its inverse operation, and $\oplus$ indicates the feature concatenation. In training phase, similar to LRN, FDN is trained with cropped regions to focus on local extraction of high-frequency signals. Particularly, to cooperate with LRN and following fusion module, the regions are cropped from the same locations in original input X_I as the regions for LRN. In testing phase, the entire image X_I is fed into FDN to acquire the complete high-resolution content layer. Thereby FDN can address the fine detail loss due to downsampling and distortions in removal process, with complementary high-resolution content in luminance channel. Therefore, integrated with the moiré-free color information from the global to local removal branch, the separated high-resolution content can contribute to a faithfully restored result with details preserved.

4.3 Layer fusion and refinement

Now that we have acquired the moiré pattern removal result X_{LR} that lacks fine details, and separated high-resolution content X_{FD}^{Y} in luminance without chrominance information, we finally fuse them to form a complete colored high-resolution output, where fine details are retained and moiré patterns are eliminated. Whereas the direct superimposition of fine details like sharp edges onto blurry layers of X_{LR} leads to artifacts like boundary shifts, we propose a light-weighted fusion and refinement network (FRN) to implement the fusion.

FRN also employs a similar stage-adaptive input strategy to conform to the regional outputs from LRN and FDN for fusion. We convert the result of LRN to YUV color space and extract the chrominance channels (U and V channels), and concatenate them with the high-resolution luminance layer X_{FD}^{Y}. As such the FRN receives and fuses a complete YUV color space representation of desired output, refines the artifacts in fusion, then finally converts the result back to RGB color space. Therefore, the complete symbolic description of our proposed pipeline can be formulated as:

$$X_O = FRN(X_{FD}^{Y} \oplus X_{LR}^{U} \oplus X_{LR}^{V}), \tag{3}$$

where X_O stands for the demoiréd output image, X_{LR}^{U} and $X_{LR}V$ for U, V channel of X_{LR} respectively.

4.4 Training loss and implementation details

Noticing that the distortions caused by equipped cameras on phones cannot be accurately calibrated pixel-wise, we adopt the Contextual Bilateral loss (CoBi loss) [38] to address such a misalignment. It matches features from source and target images to measure the similarity between unaligned image pairs. Specifically, the CoBi loss can be formulated as:

$$\mathcal{L}_{CoBi}(P,Q) = \frac{1}{N}\sum_{i}^{N} \min_{j=1,\ldots,M} (\mathbb{D}(p_i, q_j) + w_s \mathbb{D}^{'}(p_i, q_j)), \tag{4}$$

where p_i, q_j stand for the feature vectors from source image P and target image Q. N, M denote the amounts of features, $\mathbb{D}$ is the cosine distance to measure feature similarity, $\mathbb{D}^{'}$ is L2 distance between spatial coordinates, and w_s denotes the weight of spatial awareness. In the training process of each network, P is substituted with the outputs of GDN, LRN, or FRN, and Q is the corresponding ground truth images. We substitute CoBi Loss with perceptual loss [7] for FDN to suppress the artifacts emerging in the frequency to spatial transformation.

We implement the proposed framework[6] with PyTorch platform, on a PC equipped with an Intel i7-7700 3.60GHz CPU and NVIDIA 1080 Ti GPU. As for training data, we apply the FHD images in FHDMi dataset as the training input for GDN, and 384×384 regions cropped from the former images as the

[6] Detailed network architecture can be found in the supplement.

Table 2. Quantitative comparisons evaluated on different benchmarks measured by average PSNR, SSIM, and LPIPS. Larger values (↑) indicate better image quality for PSNR and SSIM, and in contrast, smaller values (↓) in LPIPS denote higher similarity to the ground truth. Red and blue denote the first and second-best method respectively.

Dataset	Method	PSNR↑	SSIM↑	LPIPS↓
FHDMi	Input	17.9740	0.7033	0.2837
	DMCNN [28]	21.5377	0.7727	0.2477
	MDDM [1]	20.8314	0.7343	0.2515
	MopNet [4]	22.7559	0.7958	0.1794
	FHDe^{2}Net	22.9300	0.7885	0.1688
TIP18 [28]	Input	20.3000	0.7380	
	MopNet [4]	27.7500	0.8950	
	FHDe^{2}Net	27.7850	0.8960	

input for LRN and FDN. Concerning parameters in training, we set the batch size at 2, initial learning rate at 0.0002, weight decay at 0.0001, and momentum at 0.9. We extracted deep feature by VGG-19 [26], and adopt conv3_2 feature for CoBi loss and conv1_2, conv2_2 feature for perceptual loss.

5 Experiments

We conduct quantitative and qualitative comparisons to evaluate the performance of FHDe^{2}Net against state-of-the-art demoiréing methods, and testify the effectiveness of each part through ablation studies. For comparison, we refer to the multi-scale learning method DMCNN [28] and MDDM [1], and channel-wise edge and binary classification guided MopNet [4][7]. The framework of these previous methods cannot directly handle the high-resolution input, because of the excessive memory occupation of their frameworks in training. Therefore, for fair comparison on high-resolution data, the methods are all trained with high-resolution regions cropped from FHDMi dataset, whose sizes are determined according to the original input size in their works. And in testing, to alleviate the boundary artifacts of patch stitching, we feed the entire high-resolution images into the retrained models, similar to the training strategy for LRN.

5.1 Quantitative evaluation

For quantitative comparison on FHDMi dataset, apart from the widely adopted metrics of PSNR and SSIM, we adopt a more recently proposed quality assessment metric LPIPS [37] for image pairs with distortions. PSNR and SSIM are purely pixel-wise metrics, and LPIPS measures perceptual image similarity

[7] According to [28, 4], the learning based methods by and large outperform traditional optimization based methods [33, 32], thus only learning methods are included.

Fig. 8. Visual quality comparison among DMCNN [28], MDDM [1], MopNet [4], and FHDe^{2}Net. Red boxes show zoom-in regions for demonstrating better details. More results are in the supplement.

using a pre-trained deep model, which evaluates the image quality beyond aligned pixels. In our case, with moderate misalignment caused by lens distortions, the pixel-wise metrics are basically fair, since most regions of the images are marginally affected by the distortions except the corners. And as a feature level perceptual metric that correlates well with human perception [37, 24], LPIPS can better handle unaligned data pairs [38] like the camera phone captured ones.

When tested with high-resolution full-screen data, it can be observed in Table 2 that FHDe^{2}Net significantly improves the visual quality of original moiré-contaminated input, and outperforms state-of-the-art methods on PSNR and LPIPS with obvious gains. Also, FHDe^{2}Net achieves the second best quantitative result by a very narrow margin on SSIM. This verifies the efficacy of FHDe^{2}Net for practical implementation of demoiréing on camera phone captured FHD images. Global-only DMCNN [27] and MDDM [1] cannot provide decent performance because of their simple learning strategy and over reliance on pixel-wise constraints. MopNet [4] delivers better results since it takes several assumptions on moiré pattern properties as learning prior, and employs feature level supervision in training. However the restriction of addressable input size within its framework design makes it only capable of learning from cropped region of high-resolution images.

To testify the generalization capability of FHDe^{2}Net framework on general moiré datasets, we fine-tune and test the GDN module of our model with existing low-resolution moiré image benchmark [28]. The results are shown in the lower

Table 3. Quantitative results of different variants of FHDe^{2}Net.

	w/o GDN	w/o LRN	w/o FDN	w/o FRN	FHDe^{2}Net
PSNR↑	18.5392	20.3143	22.9017	22.4153	22.9300
SSIM↑	0.7239	0.7652	0.7644	0.7800	0.7885
LPIPS↓	0.2577	0.1941	0.2582	0.2101	0.1688

part of Table 2, note the data in [28] are well aligned, only the pixel-wise metrics are reported. We can find that only one part of FHDe^{2}Net can still slightly surpasses the SOTA method MopNet [4]. And if we further compare the performances across datasets, it is evident that PSNR and SSIM on high-resolution data still have a large gap to those on low-resolution data. This can be attributed to the challenging nature of high-resolution data, including the misalignments of 5-10 pixel shifts on edges as shown in Figure 4 (d), which can also be inferred from the lower quantitative scores of high-resolution inputs. We have also tested on AIM [6] (LCDMoiré), the performance of FHDe^{2}Net is 41.4 on PSNR (the only metric reported on AIM online leaderboard), comparable to the second-best method (41.8) in the challenge.

5.2 Qualitative evaluation

We present the FHD qualitative comparisons against other methods in Figure 8. As we can observe, DMCNN [28] cannot thoroughly remove the moiré patterns in images, because its multi-scale learning strategy fails to catch the wide scale range of moiré patterns in high-resolution inputs. Besides, the direct superimposition of results across different scales sometimes induces block-shaped artifacts into the results, as shown in the left example of Figure 8. In contrast, FHDe^{2}Net can conserve the fine details that can only be seen with high resolution while other methods cannot, like the subtle edges of the tennis net in the left example. MDDM [1] only lightens the color of the pattern stripes, and also induces undesired blurriness to the fine details in image as shown in the first column, due to its simple learning constraint. MopNet [4] generates more visually pleasing results, yet the results also show moderate pattern residues as the moiré patterns are hard to eradicate in a single pass either globally or at patch level. On the contrary, FHDe^{2}Net effectively eliminates the patterns across different scales, including local thin steaks (middle example) and wide patterns of larger scale (right example), which is more obvious at global level.

5.3 Ablation study

In this section, we investigate the performance of different variants of proposed FHDe^{2}Net. The numerical results are presented in Table 3, where we can conclude that all functional modules contribute to a performance gain. Specifically, the global to local pattern removal modules, *i.e.*, GDN and LRN, make up

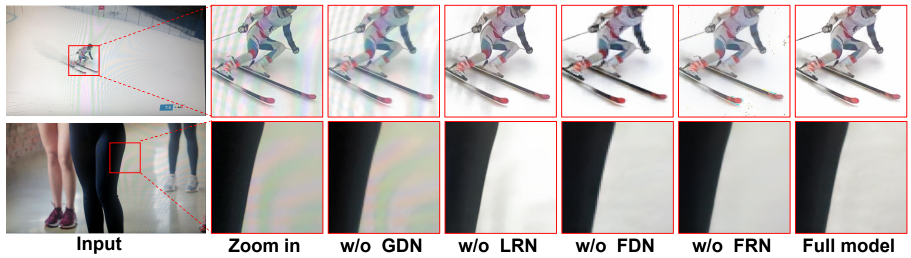

Fig. 9. Visual comparison among different variants of FHDe^{2}Net. Red boxes show zoom-in regions for better details. More results are in the supplementary material.

the major backbone of FHDe^{2}Net, as the models without GDN or LRN face a remarkable performance drop on all metrics. The high-resolution separation module FDN and fusion module FRN contribute to the enhancement of results more perceptually, as we can observe in Table 3, the lack of the two modules leads to an obvious gap to the complete model on the perceptual metric LPIPS.

Qualitative comparisons among different model variants are exhibited in Figure 9. From the zoomed-in regions, we can infer that GDN and LRN determine the existence of pattern residues. Colored stripes remain in the results by model without GDN, since such a model variant loses global perception of pattern distributions. Without the local refinement of LRN, there tend to be fragmented pattern residues in results, as shown in the top example. Deterioration of high-resolution details emerges when FDN is missing, with blurriness and jagged edges as shown in the bottom example. And FRN prevents the results from artifacts in fusion like the spots along the edges in the second shown case.

6 Conclusion

We propose a framework named FHDe^{2}Net to tackle the challenges of high-resolution image demoiréing in practical application scenarios, and provide a full high definition screen captured moiré image dataset for benchmarking this task. To the best of out knowledge, FHDe^{2}Net is the first demoiréing method capable of handling FHD images. This framework leverages a global to local pattern removal strategy, and a frequency based high-resolution content separation mechanism, to address the problems of wider pattern scale range and fine detail preservation in high-resolution images. Experimental comparisons across different datasets validate the effectiveness of FHDe^{2}Net on eradicating moiré patterns on FHD images, outperforming existing SOTA demoiréing methods.

Acknowledgments. This work was supported by National Natural Science Foundation of China under Grant No. U1611461, No. 61872012, National Key R&D Program of China (2019YFF0302902), and Beijing Academy of Artificial Intelligence (BAAI).

References

1. Cheng, X., Fu, Z., Yang, J.: Multi-scale dynamic feature encoding network for image demoiréing. arXiv preprint arXiv:1909.11947 (2019)
2. Guo, J., Chao, H.: Building dual-domain representations for compression artifacts reduction. In: Proceedings of the European Conference on Computer Vision. pp. 628–644 (2016)
3. Guo, S., Yan, Z., Zhang, K., Zuo, W., Zhang, L.: Toward convolutional blind denoising of real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1712–1722 (2019)
4. He, B., Wang, C., Shi, B., Duan, L.Y.: Mop moire patterns using mopnet. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2424–2432 (2019)
5. Huang, H., Nie, G., Zheng, Y., Fu, Y.: Image restoration from patch-based compressed sensing measurement. Neurocomputing **340**, 145–157 (2019)
6. Ignatov, A., Timofte, R., Ko, S.J., Kim, S.W., Uhm, K.H., Ji, S.W., Cho, S.J., Hong, J.P., Mei, K., Li, J., et al.: Aim 2019 challenge on raw to rgb mapping: Methods and results. In: Proceedings of the IEEE International Conference on Computer Vision. vol. 5, p. 7 (2019)
7. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European conference on computer vision. pp. 694–711. Springer (2016)
8. Kim, T.H., Park, S.I.: Deep context-aware descreening and rescreening of halftone images. ACM Transactions on Graphics **37**(4), 1–12 (2018)
9. 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)
10. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4681–4690 (2017)
11. Lefkimmiatis, S.: Universal denoising networks: a novel cnn architecture for image denoising. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3204–3213 (2018)
12. Li, B., Peng, X., Wang, Z., Xu, J., Feng, D.: Aod-net: All-in-one dehazing network. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 4770–4778 (2017)
13. Li, S., Araujo, I.B., Ren, W., Wang, Z., Tokuda, E.K., Junior, R.H., Cesar-Junior, R., Zhang, J., Guo, X., Cao, X.: Single image deraining: A comprehensive benchmark analysis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3838–3847 (2019)
14. Li, X., Wu, J., Lin, Z., Liu, H., Zha, H.: Recurrent squeeze-and-excitation context aggregation net for single image deraining. In: Proceedings of the European Conference on Computer Vision. pp. 254–269 (2018)
15. Liu, B., Shu, X., Wu, X.: Demoiréing of camera-captured screen images using deep convolutional neural network. arXiv preprint, arXiv:1804.03809 (2018)
16. Liu, D., Wen, B., Fan, Y., Loy, C.C., Huang, T.S.: Non-local recurrent network for image restoration. In: Advances in Neural Information Processing Systems. pp. 1680–1689 (2018)

17. Liu, F., Yang, J., Yue, H.: Moiré pattern removal from texture images via low-rank and sparse matrix decomposition. In: Proceedings of the IEEE Visual Communications and Image Processing. pp. 1–4 (2015)
18. Menon, D., Calvagno, G.: Color image demosaicking: An overview. Signal Processing: Image Communication **26**(8-9), 518–533 (2011)
19. Nishioka, K., Hasegawa, N., Ono, K., Tatsuno, Y.: Endoscope system provided with low-pass filter for moire removal (Feb 15 2000), US Patent 6,025,873
20. Otsu, N.: A threshold selection method from gray-level histograms. IEEE transactions on systems, man, and cybernetics **9**(1), 62–66 (1979)
21. Pekkucuksen, I., Altunbasak, Y.: Multiscale gradients-based color filter array interpolation. IEEE Transactions on Image Processing **22**(1), 157–165 (2012)
22. Plotz, T., Roth, S.: Benchmarking denoising algorithms with real photographs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1586–1595 (2017)
23. Schöberl, M., Schnurrer, W., Oberdörster, A., Fößel, S., Kaup, A.: Dimensioning of optical birefringent anti-alias filters for digital cameras. In: IEEE International Conference on Image Processing. pp. 4305–4308 (2010)
24. Shen, Z., Huang, M., Shi, J., Xue, X., Huang, T.S.: Towards instance-level image-to-image translation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3683–3692 (2019)
25. Siddiqui, H., Boutin, M., Bouman, C.A.: Hardware-friendly descreening. IEEE Transactions on Image Processing **19**(3), 746–757 (2009)
26. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
27. Sun, B., Li, S., Sun, J.: Scanned image descreening with image redundancy and adaptive filtering. IEEE Transactions on Image Processing **23**(8), 3698–3710 (2014)
28. Sun, Y., Yu, Y., Wang, W.: Moiré photo restoration using multiresolution convolutional neural networks. IEEE Transactions on Image Processing **27**(8), 4160–4172 (2018)
29. Wan, R., Shi, B., Duan, L.Y., Tan, A.H., Kot, A.C.: Benchmarking single-image reflection removal algorithms. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 3922–3930 (2017)
30. Wan, R., Shi, B., Duan, L.Y., Tan, A.H., Kot, A.C.: Crrn: Multi-scale guided concurrent reflection removal network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4777–4785 (2018)
31. Yang, C., Lu, X., Lin, Z., Shechtman, E., Wang, O., Li, H.: High-resolution image inpainting using multi-scale neural patch synthesis. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6721–6729 (2017)
32. Yang, J., Liu, F., Yue, H., Fu, X., Hou, C., Wu, F.: Textured image demoiréing via signal decomposition and guided filtering. IEEE Transactions on Image Processing **26**(7), 3528–3541 (2017)
33. Yang, J., Zhang, X., Cai, C., Li, K.: Demoiréing for screen-shot images with multi-channel layer decomposition. In: Proceedings of the IEEE Visual Communications and Image Processing. pp. 1–4 (2017)
34. Yue, H., Mao, Y., Liang, L., Xu, H., Hou, C., Yang, J.: Recaptured screen image demoiréing. IEEE Transactions on Circuits and Systems for Video Technology (2020)
35. Zeng, Y., Zhang, P., Zhang, J., Lin, Z., Lu, H.: Towards high-resolution salient object detection. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 7234–7243 (2019)

36. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 695–704 (2018)
37. Zhang, R., Isola, P., Efros, A.A., Shechtman, E., Wang, O.: The unreasonable effectiveness of deep features as a perceptual metric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 586–595 (2018)
38. Zhang, X., Chen, Q., Ng, R., Koltun, V.: Zoom to learn, learn to zoom. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3762–3770 (2019)
39. Zhang, X., Ng, R., Chen, Q.: Single image reflection separation with perceptual losses. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4786–4794 (2018)