Coarse-to-fine Disentangling Demoiréing Framework for Recaptured Screen Images

Ce Wang, Bin He, Shengsen Wu, Renjie Wan, *Member, IEEE*
Boxin Shi, *Senior Member, IEEE* and Ling-Yu Duan, *Member, IEEE*

Abstract—Removing the undesired moiré patterns from images capturing the contents displayed on screens is of increasing research interest, as the need for recording and sharing the instant information conveyed by the screens is growing. Previous demoiréing methods provide limited investigations into the formation process of moiré patterns to exploit moiré-specific priors for guiding the learning of demoiréing models. In this paper, we investigate the moiré pattern formation process from the perspective of signal aliasing, and correspondingly propose a coarse-to-fine disentangling demoiréing framework. In this framework, we first disentangle the moiré pattern layer and the clean image with alleviated ill-posedness based on the derivation of our moiré image formation model. Then we refine the demoiréing results exploiting both the frequency domain features and edge attention, considering moiré patterns' property on spectrum distribution and edge intensity revealed in our aliasing based analysis. Experiments on several datasets show that the proposed method performs favorably against state-of-the-art methods. Besides, the proposed method is validated to adapt well to different data sources and scales, especially on the high-resolution moiré images.

Index Terms—Moiré pattern, signal aliasing, layer separation.

1 Introduction

Nowadays, electronic screens have become a ubiquitous medium for conveying visual information from computers to users. To record and share information instantly, screen capture via softwares is a popular choice, but the control system of the screens is often not available to observers, thus capturing the contents presented on screens with portable cameras like smartphones is often the case in daily life. However, such recaptured screen images tend to be visually degraded by the moiré pattern artifacts as shown in Figure 1. Moiré patterns originate from the aliasing effect between the grids of the camera sensor array and screen sub-pixels in recaptured screen images, and their existence across the whole image plane severely deteriorates the image's visual quality. Therefore, screen image demoiréing, which indicates the removal of moiré patterns for recaptured screen images, is of great practical interest.

Demoiréing is challenging due to moiré patterns' complex and diversified structures, and the complicated factors influencing the pattern formation. Specifically, the pattern within an image often spans across a wide scale range with spatially-varying regional color distortions as shown in Figure 1, making it hard to describe the pattern with several typical degradation categories as in deraining [32]. Besides, different from the physical model based image restoration tasks like dehazing [16], the moiré pattern model is closely related to cameras' intrinsic and extrinsic parameters, which can hardly be explicitly expressed using an analytical model with a handful of variables. Early attempts for recaptured screen image demoiréing include both anti-aliasing optical filters [12], and post-processing based algorithms based on assumptions of moiré pattern distribution like gradient sparsity [8], [28]. However, these efforts achieve limited success because anti-aliasing low-pass filters tend to induce obvious over-smoothing artifacts, and signal processing based optimization have limited modeling capability for diversified moiré patterns.

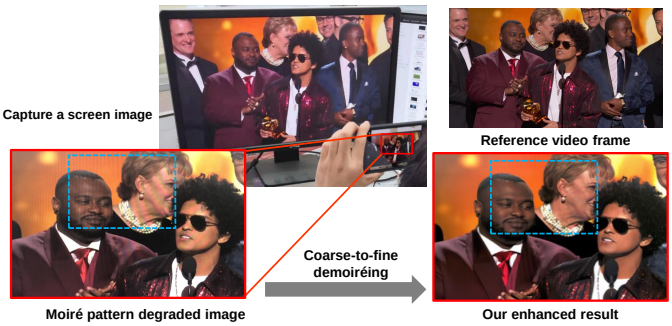


Fig. 1: A typical scenario of capturing the content on screens with smartphones. The blue dashed boxes emphasize moiré pattern's spatially-varying structure. With the proposed demoiréing framework, the moiré pattern in the recaptured screen image can be eliminated as shown on the right.

Recent methods start to exploit deep learning techniques to ease the difficulty in optimization in a data-driven man-

- *This project is supported in part by National Natural Science Foundation of China under Grant No. 62088102 and 62136001, , and in part by the PKU-NTU Joint Research Institute (JRI) sponsored by a donation from the Ng Teng Fong Charitable Foundation. Renjie Wan is supported by the Blue Sky Research Fund of HKBU under Grant No. BSRF/21-22/16 and Guangdong Basic and Applied Basic Research Foundation under Grant No. 2022A1515110692*

- *C. Wang, B. He, S. Wu, B. Shi, and L.-Y. Duan are with National Engineering Research Center of Visual Technology, School of Computer Science, Peking University, Beijing 100871, China. B. Shi and L.-Y. Duan are also with the Peng Cheng Laboratory, Shenzhen 518000, China. R. Wan is with Hong Kong Baptist University, Hong Kong SAR, China. L.-Y. Duan is the corresponding author.*

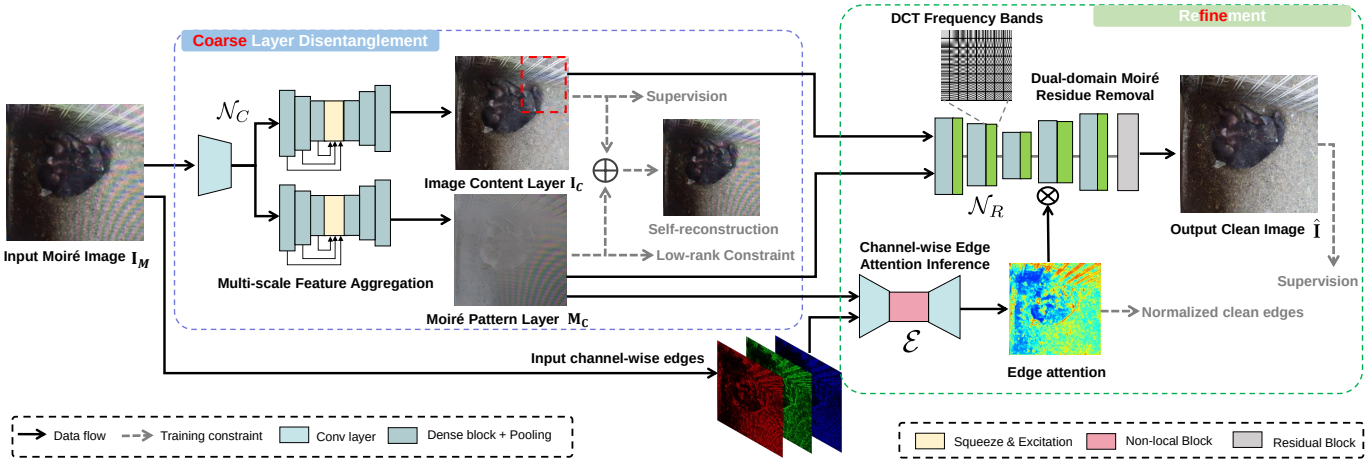


Fig. 2: Illustration of the proposed coarse-to-fine disentangling demoiréing framework, consisting of two stages: The coarse layer disentanglement stage that predicts a clean content layer and a moiré pattern layer by multi-scale feature aggregation, and the refinement stage featuring dual-domain residue removal and edge attention.

ner. Considering demoiréing as an image-to-image translation between a moiré mixture image and a corresponding clean one, end-to-end learning based demoiréing models are proposed and trained with paired images, featuring different network designs such as multi-resolution processing [22] or frequency based deep feature representations [9], [36]. Our preliminary work [3], [4] also exploit several observations on pattern appearance to design specific demoiréing modules. In [4], we designed individual modules considering the multiple pattern scale within a moiré image, the channel-wise edge difference, and pattern appearance attribute classification. In [3], we designed a frequency domain detail compensation module given the periodicity of high-frequency moiré patterns, after the downsampling based global demoiréing that reduces computational costs and enlarges the network's receptive field. Though these methods have improved demoiréing performance on benchmark datasets, their model designs are generally intuitive based on isolated observations or empirical models, without investigating the aliasing nature behind such observations or deriving an interpretable solution from an explicit moiré image formation model.

In this paper, we first look into the formation of moiré-contaminated recaptured screen images from the perspective of signal aliasing, and accordingly propose a coarse-to-fine disentangling demoiréing framework as shown in Figure 2. Specifically, we formulate the moiré image formation model as a combination of the signal aliasing effect and camera imaging process, and then conduct disentanglement based on a coarse layered model derived from the formulated model, which alleviates the ill-posedness induced by the image signal processing (ISP) pipeline. To compensate for the inaccuracy of the coarse disentanglement, we further conduct dual-domain refinement, exploiting moiré patterns' frequency domain distribution to eliminate residues and inferring edge attention to complement image details. Compared to our preliminary works [3], [4], our new framework has a different coarse-to-fine integral framework based on our moiré image formation model, and a new layer disentanglement module for coarse but less under-constrained demoiréing. Meanwhile, we also inherit the effective designs from these works, including the multi-scale feature aggregation [4], channel-wise edge cues [4] and frequency domain features [3], to build the complete framework.

The extended contributions in this paper can be summarized as follows: First, we investigate the formation model for recaptured screen moiré images based on the aliasing effect, and propose a coarse-to-fine demoiréing framework according to the derivation of the formation model. Second, we propose a new layer disentanglement module exploiting the self-reconstruction constraint and low-rank constraint, for predicting the moiré pattern layer and the clean image layer simultaneously with alleviated ill-posedness. Third, besides more comprehensive experimental comparisons, we validate the proposed framework's generalization capability across different training data scale and data sources, and discuss about the model efficiency for high-resolution inputs and moiré pattern extraction by disentanglement.

2 Related Work

As a common degradation in digital images, moiré patterns have been widely studied. In terms of the interference source in the image scene, the moiré-contaminated images can be categorized into screen moiré images and textured moiré images, originating from sub-pixel layout of LCD screens and natural fine-grained textures, respectively. The moiré patterns in textured images only mingle with regions presenting high spatial frequency. In contrast, the patterns in recaptured screen images can emerge across the entire image with more spatially-varying structures and a larger scale range regardless of the image structure shown on the screen. In this paper, we focus on the removal of moiré patterns from screens.

Early practices on demoiréing focused on improving the imaging process of cameras, including efforts on optical anti-aliasing filters [12], [15], [18] and improving the sampling of CFA via interpolation [11], [14]. However, such device-focused pre-processing methods either deteriorate image's visual quality during filtering out high-frequency components to be captured, or pose challenges to common camera image signal processing pipeline.

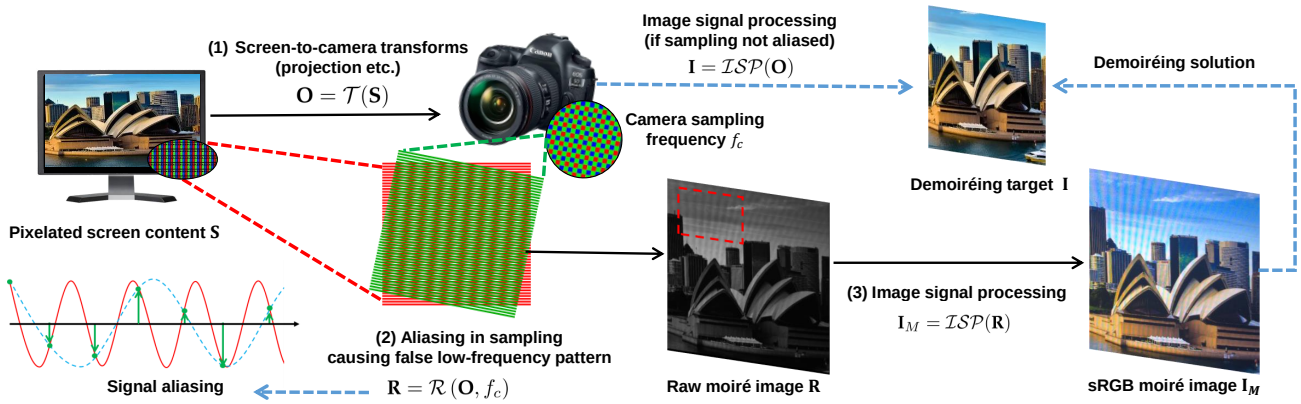


Fig. 3: The formation process of moiré images alongside the imaging pipeline. The lower left corner illustrates the false pattern in signal aliasing (Equation 1). The dashed arrows reveal the relationship between moiré image formation and the demoiréing solution.

Post-processing methods [8], [17], [19], [21], [27], [28] were proposed for removing specific type of moiré patterns like monochrome patterns or patterns in scanned images [19], [21], and assumptions on pattern characteristics were also proposed to constrain the optimization process, such as pattern sparsity and image low-rank property for textured moiré images [8], [28]. However, these signal processing based methods tend to achieve limited success when the assumptions do not hold, and suffer from high computation cost for optimization.

Deep learning has recently facilitated demoiréing methods better modeling various appearances of moiré patterns with paired training data, and such learning based demoiréing mainly focuses on the recaptured screen images because it is hard to provide moiré-free ground truth images for the textured moiré images. Sun *et al.* [22] proposed a multi-scale demoiréing network along with the first benchmark dataset that captures real LCD screens for training and evaluating demoiréing models. He *et al.* [4] made additional annotations on the data in [22] for enhancing demoiréing model. Yue *et al.* [31] proposed a multiplicative operation based network to purse simultaneous moiré pattern removal and image brightness improvement. Zheng *et al.* [36], [37] and He *et al.* [3] further exploited DCT domain priors for screen image demoiréing to boost performance, while Liu *et al.* [9] exploited the wavelet domain features to separate the frequencies of moiré patterns. Experiments also demonstrated that frequency based demoiréing can benefit the detail restoration in high-resolution screen images [3]. Recently, Liu *et al.* [10] attempted to introduce additional input of a focused-defocused image pair for demoiréing both texture and screen images with self-supervision.

Apart from those real data based methods, Liu *et al.* [7] and Yuan *et al.* [29], [30] respectively proposed various network structures for demoiréing on different synthetic data that simulate the formation of moiré patterns. These learning based methods have shown promising results for demoiréing by exploiting either spatial domain properties or frequency domain features, while our preliminary work [3] adopt a frequency based module to preserve the fine details in high-resolution images after demoiréing the downsampled image in sRGB space. In this paper, by investigating the formation process of moiré patterns alongside the in-camera imaging pipeline, we propose a coarse-to-fine disentangling demoiréing framework for screen images integrating the edge cues and the dual-domain features for refining demoiréing results.

3 Moiré Image Formation: A Deeper Look into Aliasing

Previous demoiréing works [4], [22], [36] attributed the moiré patterns in recaptured screen images to the screen-camera interference, yet more in-depth analysis on the aliasing phenomenon behind such interference is missing in their methodology, with the aliasing-related pattern properties unexplored. In this section, we investigate the in-camera formation model of moiré patterns by the clue of aliasing. The pattern formation model guides our methodology, in terms of the overall demoiréing pipeline, which is derived as the inverse process of the model, and the module components exploiting the structural and statistical pattern properties revealed by the model.

3.1 Aliasing-based moiré image analysis

Aliasing in signals. For signals, aliasing occurs when the sampling rate is not high enough to provide sufficient sampled points for faithful reconstruction. As the Nyquist's theorem reveals, only the components with frequencies below half the sampling rate can be reconstructed exactly, while frequencies above that threshold become aliased, turning into lower frequencies according to the sampling rate [13]. The aliased signal can be thereby formulated as:

$$\mathcal{R}(O, f) = O * F_{\downarrow} + \mathcal{A}(O * F_{\uparrow}, f), \tag{1}$$

where $\mathcal{R}(O, f)$ represents the reconstructed aliased signal after sampling, given the original signal O and the sampling frequency f. $F_{\downarrow}$ and $F_{\uparrow}$ denote the low-pass and high-pass filters whose cut-off frequencies are both $f/2$, dividing the signal into low-frequency components $O * F_{\downarrow}$ and high-frequency components $O * F_{\uparrow}$. Thus $\mathcal{A}(O * F_{\uparrow}, f)$ represents the false components caused by the aliasing, where one frequency is mapped to another lower one as shown in Figure

3. The mapping is a periodic function for one-dimensional signals[1], but is hard to be expanded analytically for two-dimensional images.

Aliasing based moiré image formation model. Above we only describe the aliasing for an arbitrary 1-D signal, but for moiré-degraded images, there are many more factors in its formation besides the aliasing in the sensor sampling. To comprehensively describe how the screen turns into the recaptured image, we further combine it with the imaging process.

Before sampling, the pixelated screen content $\mathbf{S}$ first goes trough a series of transforms $\mathcal{T}(\cdot)$ according to camera parameters, and turns into the actual screen content $\mathbf{O}$ arriving at the sensor array as:

$$\mathbf{O} = \mathcal{T}(\mathbf{S}). \tag{2}$$

Note that here we use bold upper case to notate the variables in matrix format, to distinguish them from 1-D signals in Equation 1. Considering the aliasing process in Equation 1, we can formulate the sensor sampling for $\mathbf{O}$ as follows:

$$\mathbf{R} = \mathcal{R}\left(\mathbf{O}, f_c\right) = \mathbf{O} * \mathbf{F}_{\downarrow} + \mathcal{A}\left(\mathbf{O} * \mathbf{F}_{\uparrow},\ f_c\right), \tag{3}$$

where $\mathbf{R}$ denotes the produced raw image, f_c denotes the sensor array's spatial frequency. Similar to Equation 1, $\mathbf{F}_{\uparrow}$ and $\mathbf{F}_{\downarrow}$ denote two-dimensional filters with Nyquist cut-off frequency $f_c/2$. As shown in Figure 3, the moiré patterns are already induced to the recaptured screen image in the raw image, and such a linear image needs to be demosaicked and adjusted for display by the image signal processing (ISP) pipeline. Thus the final moiré-contaminated sRGB image should be further expressed as:

$$\mathbf{I}_M = \mathcal{ISP}(\mathbf{R}) = \mathcal{ISP}\left(\mathbf{O} * \mathbf{F}_{\downarrow} + \mathbf{A}\right), \tag{4}$$

where $\mathbf{I}_M$ denotes the recaptured screen image with moiré patterns, $\mathcal{ISP}(\cdot)$ denotes the in-camera single processing operation, and $\mathbf{A}$ represents the initial moiré pattern term $\mathcal{A}(\mathbf{O} * \mathbf{F}_{\uparrow}, f_c)$ in Equation 3. The formation process of moiré images in Equation 2, 3, and 4 includes the pre-sampling transformation, the aliasing during sampling, and the post-sampling image adjustment, corresponding to Steps (1) to (3) in Figure 3. Different from the image model $\mathbf{I}_M = \Phi(\mathbf{I}) + N_{moire}$ in [36], which is formulated based on empirical observations, our aliasing based moiré image model involves the main factors influencing moiré pattern appearance.

Moiré pattern property analysis. With the formation process analyzed above, we can elaborate moiré patterns' properties we observed but have not explained before [4]. According to Equation 3, we can see that the aliasing process is determined by the original signal, *i.e.*, the pixelated screen $\mathbf{O}$ captured by the camera, and the camera sampling frequency f_c. Particularly, the high-frequency components of $\mathbf{O}$ mainly consist of neighboring RGB subpixels' variation and the gaps between them, as shown in Figure 3.

We first look into the structural pattern property. Considering the non-parallel projection in Equation 2, the subpixel frequency in $\mathbf{O}$ usually varies with the position on the screen due to different scaling, which further results in position-specific pattern frequency in Equation 3, visualized as the spatially-varying pattern scale as shown in Figure 1. But such frequency difference in $\mathbf{O}$ is usually not evident within neighboring regions, leading to locally periodic moiré patterns with similar patches, which makes the overall structure approximately low-rank, as shown in Figure 4. The spatially-varying structure poses challenge to our demoiréing solution, but the local periodicity can be exploited as a prior.

1. https://download.ni.com/evaluation/coretest/instrument-fundamentals-complete-guide.pdf

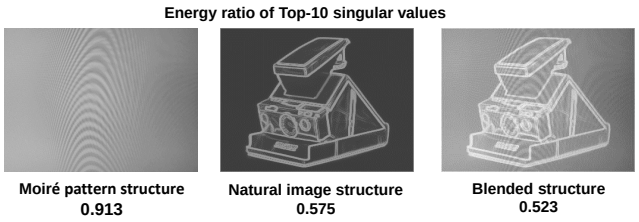


Fig. 4: The higher energy ratio of Top-10 singular values stands for higher image self-similarity, and when the ratio is above 0.7, the low-rank approximation can well describe the fine structures [24].

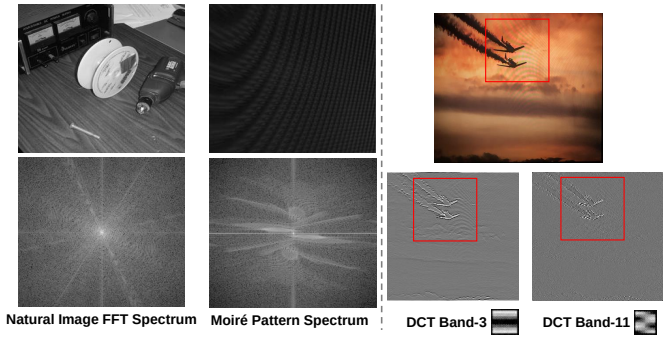


Fig. 5: Left: Comparing the spectra, we can see that the moiré pattern's energy is more centralized in non-DC components than the natural image. Right: Visualization of the 8×8 patch-wise coefficients of different DCT subbands, where the moiré pattern only shows evident response on band-3.

Besides, since the pixel-camera distance, which controls the scaling in projection, usually varies within a limited range for one image, the subpixel frequency of $\mathbf{O}$ also has a concentrated distribution. As shown in Figure 5, it also leads to the moiré pattern's relatively centralized response on certain frequency bands. Also, we can explain the channel-wise pattern intensity difference considering the camera sampling frequency f_c. In common Bayer color filter pattern, the green filters have a higher sampling frequency than the other ones, resulting in weaker aliasing on the G channel as shown in the regions marked by yellow boxes in Figure 6. These findings strengthen our understanding of the moiré pattern's nature, which favors more interpretable module designs in the following methodology section.

3.2 Demoiréing solution derivation

Intuitively, solving for the moiré-free image out of the moiré image is equivalent to disentangling the clean image and the degradation. However, considering Equation 4, the demoiréing target $\mathbf{I} = \mathcal{ISP}(\mathbf{O})$, which is the sRGB version for the unalised screen $\mathbf{O}$ captured by sensors, cannot be

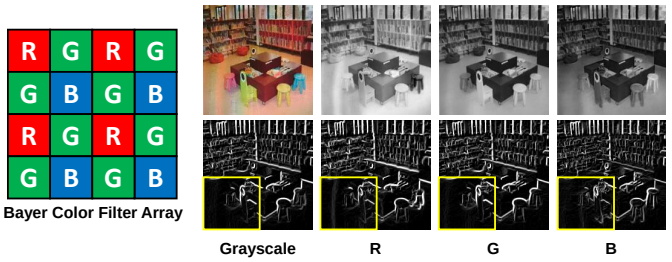


Fig. 6: The Bayer CFA and the edge maps on different sRGB color channels (better viewed when zoomed in).

naturally formulated as disentanglement owing to the existence of $\mathbf{F}_{\downarrow}$ and $\mathcal{ISP}(\cdot)$. Instead the solution of $\mathbf{I}$ should be expressed as:

$$\mathbf{I} = \mathcal{ISP}(\mathbf{F}_{\downarrow}^{-1}(\mathcal{ISP}^{-1}(\mathbf{I}_M) - \mathbf{A})), \tag{5}$$

where $\mathbf{F}_{\downarrow}^{-1}$ and $\mathcal{ISP}^{-1}$ denote the inverse mappings of $\mathbf{F}_{\downarrow}$ and $\mathcal{ISP}$. Particularly, the filter $\mathbf{F}_{\downarrow}$ here is the same as Equation 3, cutting off at half the camera sampling frequency. However, this equation is overly ill-posed to directly solve, with unknown variables $\mathbf{A}$, $\mathbf{F}_{\downarrow}$, and mapping $\mathcal{ISP}(\cdot)$. To overcome these barriers, we need to make reasonable assumptions to constrain the solution.

First, for the low-pass filtering $\mathbf{F}_{\downarrow}$, we approximately neglect it since the lost components higher than cut-off frequency, which mainly consists of the screen subpixel variation, are barely visible in recaptured screen images. Thus Equation 5 can be simplified as $\mathbf{I} = \mathcal{ISP}(\mathcal{ISP}^{-1}(\mathbf{I}_M) - \mathbf{A})$.

As for the ISP function $\mathcal{ISP}(\cdot)$, it would be ideal to solve Equation 5 by making the ISP function and its inverse mapping cancel out each other, and modifying the equation as $\mathbf{I} = \mathbf{I}_M - \mathcal{ISP}(\mathbf{A})$. However, this transformation is precisely equivalent only when the ISP function is linear. To compensate for the non-linearity in the ISP while still enjoying such a simplified formulation, we approximate the linearly-expanded equation with an offset term Δ, which denotes the pixel value deviations caused by tone-mapping. Equation 5 is then simplified as:

$$\mathbf{I} = \mathbf{I}_M - \mathbf{M} + \Delta, \tag{6}$$

where $\mathbf{M}$ denotes the term $\mathcal{ISP}(\mathbf{A})$ as an assumed moiré pattern layer. Note that Δ is a pixel-wise term with sufficiently high degrees of freedom for approximating Equation 5. Now we can directly involve our input $\mathbf{I}_M$ in the solution by disentanglement.

4 Coarse-to-fine Disentangling Demoiréing Framework

In this section, in light of the layered moiré image model derived in Section 3.2, we propose to build a disentanglement based demoiréing framework. Details about how to realize such a framework, *i.e.*, the module designs and optimization constraints, are introduced based on the pattern properties we analyzed in Section 3.1.

4.1 Framework design overview

From Equation 6, we can infer that the moiré image input $\mathbf{I}_M$ can be divided into different layers: the clean image layer $\mathbf{I}$, the moiré pattern layer $\mathbf{M}$, and the pixel offset Δ, as $\mathbf{I}_M = \mathbf{I} + \mathbf{M} - \Delta$. Among these layers, the estimation of $\mathbf{I}$ can be constrained in a supervised learning manner, while the other two remain uncertainty in optimization.

To alleviate such ill-posedness in in the disentanglement of aforementioned layers, we propose a coarse-to-fine demoiréing framework, where we first roughly disentangle a clean image layer $\mathbf{I}_C$ and a moiré pattern layer $\mathbf{M}_C$ regardless of the offset term, following a coarse model:

$$\mathbf{I}_M = \mathbf{I}_C + \mathbf{M}_C. \tag{7}$$

The subscript C in $\mathbf{M}_C$ refers to a coarse estimation of the layer $\mathbf{M}$. In the coarse disentanglement, we propose to constrain the unknown $\mathbf{M}_C$ with a low-rank constraint, considering the local periodicity pattern structural property analyzed in Section 3.1. Thereby this coarse disentanglement can be reasonably constrained since the two layers are either supervised or regularized in prediction.

Then, we refine the coarse demoiréing result $\mathbf{I}_C$ by further eliminating remaining moiré residues caused by the inaccurate layer division, guided by the implicit moiré pattern localization in predicted $\mathbf{M}_C$. Particularly, the refinement is also conducted in the frequency domain, considering moiré patterns' centralized spectrum distribution analyzed in Section 3.1.Besides, such refinement is also necessary to compensate for the neglection of the offset term as well as the image details lost in the disentanglement process, where an edge-specific attention is exploited based on the channel-wise pattern intensity difference analyzed in Section 3.1.

To summarize, we realize a coarse-to-fine dual-domain demoiréing framework as follows:

$$\{\mathbf{I}_C,\ \mathbf{M}_C\} = \mathcal{N}_C(\mathbf{I}_M), \tag{8}$$

$$\hat{\mathbf{I}} = \mathcal{N}_R(\mathbf{M}_C,\ \mathbf{I}_C,\ \mathcal{E}(\mathbf{I}_M,\ \mathbf{M}_C)). \tag{9}$$

The two equations above describe the coarse layer disentanglement stage, and the dual-domain refinement stage, respectively. $\mathcal{N}_C$ and $\mathcal{N}_R$ denote the corresponding modules. $\hat{\mathbf{I}}$ denotes the final moiré-free output, and $\mathcal{E}(\mathbf{I}_M, \mathbf{M}_C)$ stands for the edge attention extracted from the original input and the estimated coarse moiré pattern layer.

4.2 Coarse self-reconstructing layer disentanglement

As demonstrated in Section 4.1, we aim to first roughly disentangle the input moiré image as two layers, which sacrifices the precision for reducing ill-posedness in optimization. To realize such layer separation, we propose to simultaneously predict the coarse clean image layer $\mathbf{I}_C$ and the moiré pattern layer $\mathbf{M}_C$ with a two-branch multi-scale layer disentanglement module $\mathcal{N}_C$ as shown in Figure 2.

Multi-scale feature aggregation structure. The multi-branch structure for separating different image layers has been proved effective in many image-related tasks such as reflection removal [25] and intrinsic image decomposition [6]. For the demoiréing case, considering the spatially-varying scale of the moiré pattern within a single image, we propose to aggregate multi-scale spatial features to distinguish moiré patterns from image contents at different scales in the layer disentanglement.

Specifically, the disentanglement module consists of a shared feature extractor, and two branches to predict the im-

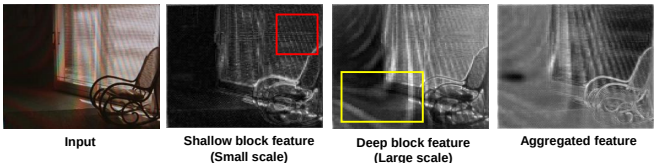


Fig. 7: Visualization of the multi-scale feature aggregation in the moiré pattern layer disentanglement branch.

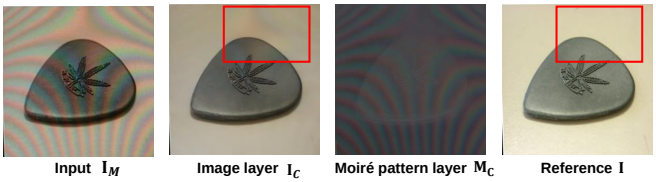


Fig. 8: An example of disentanglement results. Red boxes mark the regions containing moiré pattern residues.

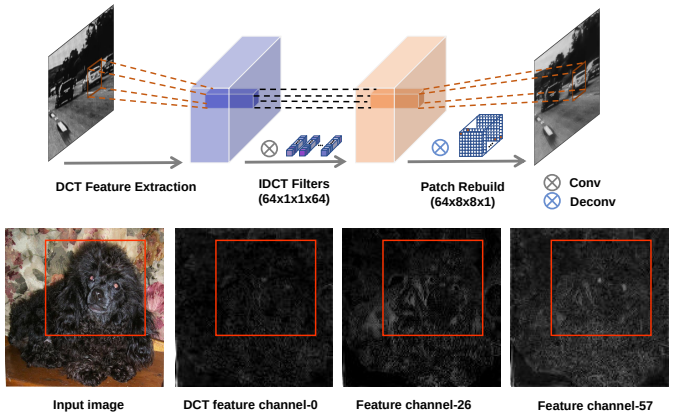


Fig. 9: Top: Illustration of the convolutional IDCT layer. Bottom: Visualization of different DCT domain feature channels before the last convolutional IDCT layer.

age content layer and the moiré pattern layer, respectively. The feature extractor takes the input recaptured screen image $\mathbf{I}_M$, and extracts feature maps from the mixture with cascaded 3×3 convolutional layers. The two layer prediction branches then take the mixture features as common input, with identical U-Net backbone structure stacked by dense blocks. Particularly, we make each dense block work on different spatial scales with pooling layers between different blocks, and further aggregate these multi-scale features by resizing and concatenating the feature maps for Squeeze-and-Excitation (SE) operation [5]. The SE block re-weights the features extracted from different scales to emphasize the dominant pattern structures. As shown in Figure 7, after the SE operation [5] in the moiré pattern layer prediction branch, both the response of high-frequency pattern structures from the shallow dense block and the response of low-frequency structures in deeper block are emphasized in the aggregated feature. Therefore, such a multi-scale feature aggregation in the layer prediction branch helps to model the moiré distortions with spatially-varying scales.

To properly optimize the learning of the coarse layer disentanglement, we need to constrain both of the learning targets, *i.e.*, the coarse clean image layer $\mathbf{I}_C$ and the moiré pattern layer $\mathbf{M}_C$. The prediction for $\mathbf{I}_C$ can be easily supervised as we can get the moiré-free ground truth screen content $\mathbf{I}$ in the training pairs.

Low-rank constraint. Though we cannot directly obtain such supervision for the moiré pattern layer, we can regularize the prediction with the low-rank assumption on pattern structure as analyzed in Section 3.2 and 4.1. Specifically, we assume the predicted moiré pattern layer to be low-rank since moiré patterns tend to present similar periodic structures within local regions, resulting in high internal correlation within the pattern. The remaining screen contents in $\mathbf{M}_C$ would increase the low-rank penalty. As shown in Figure 4, a recaptured screen image with pure white screen content, which presents no other structures except the moiré pattern, has high self-similarity. But when additional edges from another image are added, the image rank obviously increases, which verifies the rationality of low-rank constraint for $\mathbf{M}_C$. We adopt the weighted nuclear norm [26] to measure such low-rankness, which sums singular values of the matrix stacked by column vectors of image patches.

Self-reconstruction constraint. To further avoid potential trivial solutions of $\mathbf{M}_C$ constrained by the relatively loose low-rank prior only, we impose an addition constraint based on the coarse model in Equation 7, by re-combining the disentangled layers to reconstruct the input image $\mathbf{I}_M$. Such a self-reconstruction constraint is reverse to the disentanglement process, and can be formulated as: $||(\mathbf{I}_C+\mathbf{M}_C)-\mathbf{I}_M||_2$, which indirectly constraint the moiré pattern layer $\mathbf{M}_C$ given that $\mathbf{I}_C$ is fully supervised.

To sum up, the overall loss $\mathcal{L}_C$ for the coarse layer disentanglement module can be expressed as follows:

$$\mathcal{L}_C = ||\mathbf{I}_C-\mathbf{I}||_2+\lambda||g(\mathbf{M}_C)||_N+||(\mathbf{I}_C+\mathbf{M}_C)-\mathbf{I}_M||_2, \quad (10)$$

where the three terms correspond to the supervision on the clean image layer, the low-rank regularization on the moiré pattern layer, and the self-reconstruction constraint, respectively. $||\cdot||_N$ denotes the weighted nuclear norm, $g(\mathbf{M}_C)$ represents the matrix stacked by grayscale moiré pattern layer patches that emphasize on pattern structure, and λ denotes the loss weight set as 0.05. The results of both branches are involved in the self-reconstruction, also making the two branches better collaborate in disentanglement.

With the proposed coarse layer disentanglement module, we can obtain a coarse estimation of the clean image and an explicit description of the moiré pattern appearance with reduced ill-posedness. As shown in Figure 8, the moiré pattern in $\mathbf{I}_C$ is largely suppressed, yet with some pattern silhouettes remaining, which demands further refinement.

4.3 Dual-domain refinement with edge attention

The refinement after the coarse layer disentanglement makes the proposed framework work in a coarse-to-fine manner. Such refinement is necessary since the disentanglement is inaccurate following a coarse image composition model, and we aim to improve the coarse result from two major aspects: eliminating moiré pattern residues in the predicted clean image, and enhancing image details. To realize that, we conduct moiré residue removal as the backbone of the refinement stage as shown in Figure 2, where the predicted moiré pattern layer is exploited as a

crucial guidance as formulated in Equation 9, and introduce the edge attention for image details.

4.3.1 Guided dual-domain moiré residue removal

In the refinement stage, different from the disentanglement, we need to focus on the moiré pattern residue region instead of the entire image. The estimated moiré pattern layer $\mathbf{M}_C$ can serve as an ideal guide for such refinement, since it reveals the pattern's spatial and frequency characteristics on the whole. Therefore, we propose to feed the refinement module $\mathcal{N}_R$ with the coarse demoiréing result $\mathbf{I}_C$, and concatenate the moiré pattern layer $\mathbf{M}_C$ to the inputs as the additional guidance as expressed in Equation 9.

To better capture the moiré residues, we propose to build a Discrete Cosine Transform (DCT) based dual-domain network for the refinement module. The reason for exploiting frequency domain features lies in that moiré patterns tend to have more centralized responses on certain frequency bands as shown in Figure 5. Thanks to such a distribution property, the spectrum of disentangled moiré pattern layer $\mathbf{M}_C$ can shed light on the potential frequency bands that might contain moiré residues. To localize the residues simultaneously in both the spatial and frequency domains, we realize dual-domain collaboration among patch-wise DCT features that preserves the spatial relationships between patches.

As for the specific network structure, we build a dense block based U-Net featuring convolutional inverse DCT (IDCT) layers [35]. The convolutional IDCT layer is composed of 64×64 filters with 1×1 kernel, whose weights are fixed as the 64×64 IDCT matrix elements. These kernels transform the frequency components from 64 bands to 64 pixels within an 8×8 image patch, as shown in Figure 9. Such an IDCT layer enforces the features that it takes to represent an image patch with DCT coefficients, as visualized in Figure 9, there shows a correspondence between feature channels and frequency bands of the image. Specifically, we can observe that channel-26 focuses on the relatively smooth variations, and channel-57 mainly captures the high-frequency textures over the entire image. It should be noted that compared to the explicit feature domain transform in our preliminary work [3], which demands an additional DCT translation beforehand [2], the adopted implicit transform with IDCT layer can better boost demoiréing performance. We suppose this advantage benefits from higher degrees of freedom for the filters to learn frequency domain priors in terms of eliminating moiré residues.

4.3.2 Channel-wise edge attention inference

Besides the moiré pattern residues, another issue we aim to address in refinement is the edge detail loss in coarse disentanglement. As shown in Figure 8, the moiré pattern layer $\mathbf{M}_C$ wrongly contains some contours of the screen content, and the edges of its counterpart $\mathbf{I}_C$ are correspondingly weakened. To compensate for such edge deterioration, we propose to introduce an edge attention in the refinement, exploiting the channel-wise moiré pattern intensity difference as shown in Figure 6.

The attention $\mathcal{E}(\mathbf{I}_M, \mathbf{M}_C)$ is learned from original input $\mathbf{I}_M$ and the predicted moiré pattern layer $\mathbf{M}_C$, as notated in Equation 9, using $\mathbf{M}_C$ to distinguish the image details from moiré stripes in the original moiré image. We enforce the attention to focus on moiré-free edges in the spatial features, by supervising the attention inference with normalized ground truth edges.

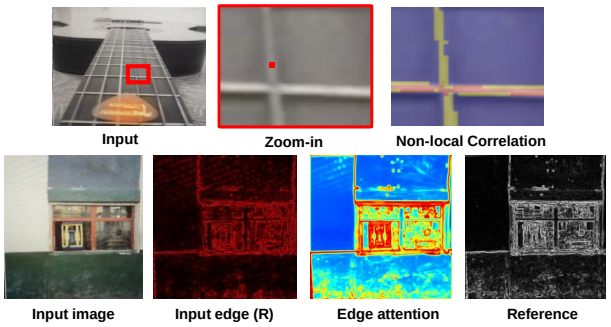


Fig. 10: Top: Non-local block helps the weak edge (red point) obtain a stronger response by strengthening its correlations with other edges. Bottom: Visualization of the predicted edge attention extracted from channel-wise input edge maps, supervised by normalized clean edges.

Specifically, the channel-wise attention inference first extracts separate edge maps in R, G, and B channels from the inputs by a fixed convolution layer with Sobel kernel, instead of the one-channel grayscale edge map. The difference among the channels can potentially reveal the moiré pattern regions, where the edge intensity in the G channel is weaker than the other two. The channel-wise features are then passed through several non-local blocks, which re-weight the feature maps based on the correlation between the feature response at one position and all other positions. Such an operation can help strengthen weak edges' long-distance dependency on other stronger edges as shown in Figure 10. Also, the predicted edge attention can emphasize the regions where the image edges are dense while filtering out most of the moiré pattern edges, and even strengthen some blurred edges in the input. It can collaborate with the dual-domain refinement well as the edges can also be interpreted as high-frequency image components.

4.3.3 Refinement process

As shown in Figure 2, the dual-domain residue removal constitutes the backbone of the refinement process, and the edge attention is used to re-weight the features on the decoder side of the IDCT-based U-Net. Then, after removing residues and emphasizing edges, we attempt to additionally adjust the tonemap offset with a residual block. We expect that the residual term predicted from the refined result can compensate for the ignored offset term Δ in Equation 6 to some extent. The final demoiréing result $\hat{\mathbf{I}}$ is then output by the residual learning.

The learning constraint in the refinement stage includes the optimization objective on the final result and the supervision of the attention learning, and can be summarized as:

$$\mathcal{L}_R = ||\mathbf{I}-\hat{\mathbf{I}}||_1+\sigma||\mathcal{V}(\hat{\mathbf{I}})-\mathcal{V}(\mathbf{I})||_2+||\mathcal{E}(\mathbf{I}_M,\mathbf{M}_C)-\mathbf{E}_\mathbf{I}||_1, \quad (11)$$

where the first two terms impose supervision on whole refinement process, and the last term is meant for the edge attention learning. Particularly, $\mathcal{V}(\cdot)$ denotes feature extraction by the ImageNet [1] pre-trained VGG-19 model

TABLE 1: Quantitative comparison on TIP-2018 dataset [22] adopting the training setup in [4]. Larger values (↑) indicate better image quality for PSNR and SSIM, and smaller values (↓) denote higher similarity for LPIPS. Red and blue denote the best and second-best methods respectively.

	PSNR↑	SSIM↑	LPIPS↓
Input	19.461	0.726	0.256
AMNet [31]	25.473	0.833	0.171
DMCNN [22]	26.101	0.844	0.125
WDNet [9]	27.120	0.854	0.130
MopNet [4]	27.483	0.861	0.093
FHD^{2}eNet [3]	27.794	0.867	0.103
MBCNN [36]	28.401	0.871	0.109
Ours	28.867	0.894	0.088

[20] for the perceptual loss whose coefficient σ is set as 0.1, and $\mathbf{E}_\mathbf{I}$ denotes the normalized edges of the ground truth $\mathbf{I}$.

4.4 Implementation details

Combining Equation 10 and 11, the complete loss function for our coarse-to-fine dual-domain demoiréing framework can be expressed as follows:

$$\mathcal{L} = \mathcal{L}_C + \mathcal{L}_R. \tag{12}$$

Particularly, we empirically set the coefficients for all loss terms except for the low-rank constraint and perceptual loss as 1, after comparing model performances during training among a few sets of weight values.

We implement the proposed framework with PyTorch on an NVIDIA 1080 Ti GPU. The standard training of the entire framework is progressively conducted with the Adam optimizer and images sized as 256×256. We train the framework for 200 epochs, and empirically set the batch size at 2, initial learning rate at 0.0002 with linear decrease, weight decay at 0.0001, and momentum at 0.9.

5 Experiments

5.1 Comparison with the state-of-the-arts

In this section, to evaluate the performance of our method, we conduct quantitative and qualitative comparisons against state-of-the-art demoiréing methods on the benchmark screen image demoiréing dataset TIP-2018 [22]. The TIP-2018 dataset consists of 135,000 screen-shot images with moire artifacts collected from the ImageNet dataset [1], with the original image shown on the screen serving as the moiré-free ground truth. The training data pairs are aligned using homography registration based on image corners.

We compare the proposed coarse-to-fine disentangling demoiréing framework with our preliminary works MopNet [4] and FHDe^{2}Net [3], and previous methods including multi-scale model DMCNN [22], multiplicative operation based network AMNet [31], wavelet based mehotd WDNet [9], and learnable bandpass filter method MBCNN [36][2].

2. We report MBCNN's performance using the model in [36] as the codes for [37] have not been released. The multiple-image based method FDN [10] is not included for fair comparison among single-image methods.

TABLE 2: Qualitative model performance comparisons when DMCNN [22] (0.3M parameters), MBCNN [36] (14.9M parameters), and the proposed model (15.4M parameters) are trained with data of different quantities and tested on the same testing set.

		PSNR↑	SSIM↑	LPIPS↓
TIP-100%	DMCNN [22]	26.770	0.871	0.112
	MBCNN [35]	30.030	0.893	0.087
	Ours	30.939	0.914	0.060
TIP-10%	DMCNN [22]	26.164	0.850	0.121
	MBCNN [35]	28.406	0.877	0.109
	Ours	28.869	0.892	0.087
TIP-5%	DMCNN [22]	22.241	0.711	0.127
	MBCNN [35]	23.405	0.781	0.116
	Ours	24.211	0.832	0.095
TIP-3%	DMCNN [22]	20.701	0.653	0.130
	MBCNN [35]	22.585	0.761	0.123
	Ours	23.598	0.814	0.107

5.1.1 *Quantitative comparison*

For quantitative evaluation, apart from the widely used measurements PSNR and SSIM in image restoration tasks, we also adopt a perceptual image similarity metric LPIPS [33], which correlates well with human perception and evaluates the image quality using a pre-trained deep model.

The quantitative comparison results are shown in Table 1. It should be noted that all the models in the table are retrained with a 10% subset (12000 out of 121500 pairs) of the complete TIP-2018 training set [22] as our preliminary work MopNet [4], which adopts such a setup due to the limit of available attribute labels. We follow this setup to make a fair comparison among all methods and reduce the time cost of training model considering the relatively large data scale of TIP-2018 [22].

From the results, we can first notice that the PSNR and SSIM values of the inputs are at a low level, which demonstrates that moiré patterns severely affect the visual quality of recaptured screen images. The demoiréing methods evidently enhance the moiré images quantitatively. Particularly, MBCNN [36] achieves the second-best place in the comparison with a frequency based network, validating the effectiveness of the frequency based priors for demoiréing. Also, we can find that the proposed coarse-to-fine framework consistently outperforms all other methods in terms of pixel-wise similarity (PSNR), global appearance indices (SSIM), and visual perception (LPIPS). The performance gains compared to our preliminary methods [3], [4] also verify the contributions of our extensions on network designs.

Impact of training data quantity. As demonstrated above, the proposed framework can perform favorably with limited training data. Considering the high costs of collecting large-scale data for adapting the model to specific application scenarios, we further investigate the influence of cutting down on training data scale on model performance. For experimental setup, besides the TIP-10% training set (12K images) adopted in the comparison across all methods, we investigate the performance of several methods on the full-

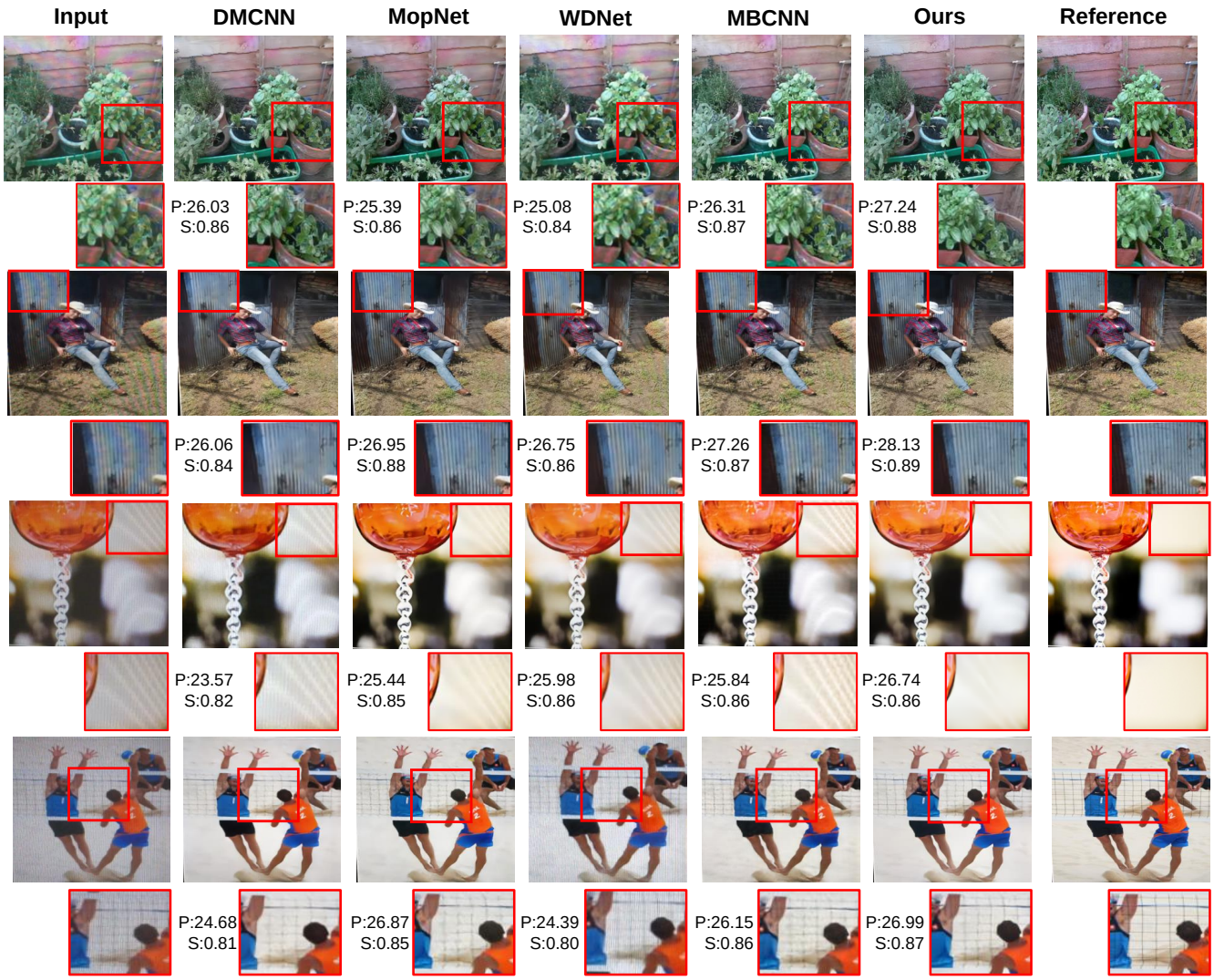


Fig. 11: Examples of demoiréing results on TIP-2018 dataset [22]. The corresponding PSNR and SSIM values are displayed below each image. Red boxes mark the regions with noticeable differences, please zoom in to view details.

size training set, the 5% quantity (6K images), and the 3% quantity (3K images) from the TIP-2018 training set. To avoid potential bias in sampling the subsets, we conduct image reshuffling and random sampling for 6 times for each data quantity, and report the average model performances in Table 2. The rationality of the sampling based experiment setup can be guaranteed, as the average TIP-10% results in Table 2 are consistent with those in Table 1 that follow the training setup of [4]. We also compare the model performances to further validate that the subset difference caused by random sampling show no evident effects on our model, which is presented in the supplementary material.

From the results in Table 2, we can observe that the quantity of training data undoubtedly affects the performance of the trained models with a positive correlation. The performance of the proposed framework is consistently superior to the previous state-of-the-art method MBCNN [36] and DMCNN [22] regardless of different training data scales. Particularly, when the training data quantity is reduced from 5% (6K) to 3% (3K), our method can still restrict the performance drop to be small, indicating our method's potential to work on scenarios where the annotated training data is much less than the testing data. Our model also has a higher performance upper bound than MBCNN [36] that has a similar model size to ours, which implies the effectiveness of our module designs. On the contrary, DMCNN's performance does not see a leap when the training data scale is largely expanded from 12K to 120K, while showing a drastic drop when the data quantity further decreases below 12k, even inferior to the original moiré image inputs. We assume that DMCNN's limited success in dealing with the variation of training data quantity might be attributed to its weaker data modeling capability caused by its lightweight and simple network structure.

5.1.2 *Visual quality comparison*

We present the qualitative comparisons against other methods in Figure 11. As we can observe, the input images are contaminated with obvious moiré patterns with various appearances, and all competing learning based demoiréing methods are able to suppress moiré patterns to some extent. Our proposed method provides better results in terms of eliminating diversely shaped moiré patterns, and maintaining edge details in image contents.

Fig. 12: Top: Visual comparison among images from TIP-2018 dataset [22] and the high-resolution moiré image dataset FHDMi [3]. Bottom: Visual comparisons among different demoiréing results for regions marked by the blue boxes.

Specifically, from the top two rows in Figure 11, we can observe that previous methods fail to thoroughly wipe out the strong color distortion of the moiré stripes in the red boxes, while our method more successfully suppresses the color distortion, which might benefit from the disentanglement of the moiré pattern layer. In the third row of Figure 11, we can observe that when confronted with the spatially-varying pattern, most methods tend to only remove the high-frequency lines, leaving obvious low-frequency pattern residues as marked by the red box. The proposed coarse-to-fine framework can better handle such cases with the dual-domain refinement stage. And in terms of preserving image details, the bottom row shows a case with subtle edges on the volleyball net as marked by the red box, the proposed method effectively alleviates the blur to the edges in demoiréing process with the edge attention, compared to other competing methods.

5.2 Adaptation for high-resolution images

The growing resolution of camera phones raises new challenges for demoiréing. On the one hand, the high image resolution expands the range of pattern scales, which can exceed the entire receptive field of common networks (around 100×100), and demands high computing costs. On the other hand, inputs with higher resolution can capture more subtle details, which are subject to over-smoothing during demoiréing. The naive solution of directly downsampling the inputs for demoiréing and scaling it back to the original resolution will inevitably cause the loss of image details.

To address the above issues, following our previous work [3], we can adapt the proposed coarse-to-fine demoiréing framework for high-resolution inputs, by training it module by module with different resolutions. Specifically, we first downsample the high-resolution moiré image to a fixed size of 384×384 before feeding it into the coarse disentanglement module. Thereby we can greatly enlarge the network receptive field on the original-sized image to globally capture large-scale moiré color distortions while keeping the computational burden low. Then, to focus on restoring the image details lost in downsampling, we train the refinement module with cropped local patches from high-resolution inputs. Particularly, to cooperate with the refinement stage, the clean image layer predicted from the rescaled input is upsampled back to the original size. In the testing phase, the coarse disentanglement module still takes the downsampled version of a high-resolution moiré image as the input, while the refinement module takes the entire high-resolution input to realize the efficient coarse-to-fine adaptation. In addition, to alleviate the nonlinear lens distortions issue in high-resolution training pairs, we also adopt the contextual bilateral loss [34], which conducts feature vector matching between the prediction and the target image, to tolerate the inevitable misalignment.

Full high definition moiré image dataset. To validate the effectiveness of such an adaptation, we conduct experiments on Full High Definition Moiré Image Dataset named FHDMi, which is collected in our preliminary work [3]. Different from the TIP-2018 dataset [22] that only crops $384 \times 384 \sim 700 \times 700$ regions from the screen, the images in FHDMi cover the whole screen with FHD resolution (1920×1080) as shown in Figure 12. Besides, the 12000 moiré images in FHDMi not only cover natural images, but also screen-specific scenarios like webpages or documents.

Experimental comparisons. Most existing demoiréing methods cannot directly handle high-resolution inputs due

TABLE 3: Quantitative comparison and model efficiency analysis on FHDMi dataset [3]. Red and blue denote the first and second-best method, respectively. The inference time and peak memory are measured for processing a single FHD resolution image. The prefix ◇ denotes the patch-based processing, which causes significant processing time increase.

	PSNR↑	SSIM↑	LPIPS↓	Parameters	Memory	Inference time
Input	17.974	0.703	0.284	N. A.	N. A.	N. A.
DMCNN [22]	21.538	0.773	0.248	0.3M	2.9 GB	0.015s
◇ MopNet [4]	22.756	0.796	0.179	58.6M	1.7 GB	10.780s
FHD^{2}e [3]	22.930	0.789	0.169	13.6M	5.9 GB	0.040s
◇ MBCNN [36]	22.847	0.805	0.203	14.9M	1.6 GB	3.800s
Ours	23.627	0.804	0.161	15.4M	5.9 GB	0.025s

Fig. 13: Visual result comparisons on different data sources: Synthetic screen moiré images [29] (top left), in-the-wild textured image [27] (top right), and moiré images with obvious color shift [31] .

to excessive memory occupation. Therefore, for a fair comparison, such methods [4], [22], [36] are all re-trained with high-resolution regions cropped from the FHDMi dataset, whose sizes are determined according to the input size used in their works. In testing, MopNet [4] and MBCNN [36] are fed with patch inputs whose demoiréing results are stitched back together because their memory demand to process a high-resolution image in the inference stage is still larger than an NVIDIA 1080Ti GPU.

The quantitative comparisons on the FHDMi dataset [3] are presented in Table 3. Note that 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 [34]. First, comparing the input column of Table 1 and Table 3 as a whole, we can infer that the high-resolution recaptured screen images are more challenging to restore than the smaller ones in TIP-2018 [22], considering the lower demoiréing quantitative performances in Table 3.

The proposed method obtains the best performance on PSNR and LPIPS, and also achieves the second best result by a very narrow margin to MBCNN [36] on SSIM. These results validate that our framework can be well adapted for the application of high-resolution image demoiréing. The reason for MBCNN's decent high SSIM value is that, to concentrate on the comparison of demoiréing and reduce the influence of the boundary artifacts in stitching patches for the patch based methods we compute PSNR and SSIM for patch based results by averaging the patch-wise indices.

The visual comparison on the FHDMi dataset is shown in Figure 12, where we can observe that DMCNN [22] cannot adapt well to high-resolution inputs, leaving intense moiré residues in the bottom row. MBCNN [36] suffers from both pattern residues and boundary artifacts derived from the patch-based processing. FHD^{2}eNet can largely suppress moiré patterns, but still with minor residues in textureless regions like the neck and the phone screen. Our method can realize more thorough elimination thanks to the improved coarse-to-fine framework structure.

Model efficiency for high-resolution inputs. Apart from the demoiréing performance, the model efficiency is also a big concern in real applications. To analyze the demoiréing models' efficiency when dealing with the FHD inputs, we also report their model parameter number, reference time, and the memory occupation for processing one image from the FHDMi dataset [3], and the results are shown in Table 3. We can find that DMCNN [22] has the highest operating efficiency due to its light-weighted structure. However, its demoiréing performance is obviously inferior to other methods as demonstrated above. MopNet [4] and MBCNN [36] compensate for its memory overrun with patch partitioning, leading to a remarkable inference time increase (over $90\times$) that degenerates the model applicability. Particularly, their patch based demoiréing results also suffer from noticeable boundary artifacts. The proposed coarse-to-fine framework and FHD^{2}eNet [3] realize a more reasonable trade-off among the demoiréing performance, memory occupation, and inference time, and keep the peak memory within the capacity of one NVIDIA 1080Ti GPU with improved model performance compared to DMCNN [22]. Thus we suppose that the adaptation for high-resolution inputs featuring image rescaling and detail compensation is beneficial for making the model more efficient in real applications.

5.3 Generalization capability analysis

To further evaluate our model's generalization potential, we investigate the proposed coarse-to-fine framework on different data sources, including synthetic data, in-the-wild textured moiré image, and recaptured screen images with obvious color shifts.

Fine-tuning results on synthetic data. We fine-tune our model on synthetic demoiréing dataset LCDMoire [29] that simulates moiré patterns from screens for generating moiré images. Since the authors only release the testing inputs without corresponding ground truth, we only present the PSNR value of our demoiréing results on such synthetic data, as measured by their demoiréing challenge website[3].

3. https://data.vision.ee.ethz.ch/cvl/aim19/

TABLE 4: Model generalization analysis on different data sources (measured by PSNR on [29], and PSNR/SSIM on [31].

	Synthetic [29]	Cross-domain [31]
DMCNN [22]	35.48	16.71/0.79
MBCNN [36]	44.04	17.01/0.75
Ours	**44.10**	**18.33/0.82**

TABLE 5: Quantitative results of different model variants trained on TIP-2018-10% subset [22]. DE, FR, EA, SRC, and LRC stand for the proposed disentanglement for the moiré pattern layer, the frequency domain based moiré residue removal, the edge attention, the self-reconstruction constraint, and the low-rank constraint, respectively.

	PSNR↑	SSIM ↑	LPIPS↓	Time
w/o DE	27.815	0.867	0.102	14.9ms
w/o FR	28.387	0.878	0.098	11.5ms
w/o EA	28.594	0.882	0.093	10.7ms
w/o SRC	27.925	0.871	0.100	16.0ms
w/o LRC	28.628	0.879	0.090	16.0ms
Full model	**28.867**	**0.894**	**0.088**	16.0ms

Our methods shows comparable performance with the state-of-the-art MBCNN [36], as presented in Table 4 ("Synthetic"). Particularly, the high PSNR values imply that the synthetically generated moiré patterns are easier to capture and remove than the real-world ones. Visual comparison in Figure 13 also verifies that our fine-tuned model can adapt well to the synthetic data and even better suppress moiré color distortions than MBCNN in some cases.

In-the-wild textured moiré images. We also test our model's generalization capability on moiré images captured in uncontrolled conditions from [27], where moiré patterns are mainly derived from highly textured surfaces like clothes. Since in-the-wild images are usually not paired with a moiré-free reference image, we only present the qualitative results as shown in Figure 13 (top right), using the pre-trained model of our method and MBCNN [36] on TIP-2018 dataset [22]. The visual comparison between two learning based demoiréing methods shows that the proposed framework can effectively suppress the moiré pattern on the bag's texture, demonstrating our method's better generalization for in-the-wild textured moiré images.

Cross-domain test against color shifts. Due to different display settings and various camera color retouching implementations, recaptured screen images that are captured with different combinations of display devices and cameras often present different degrees of color shifts compared to the original moiré-free images. Such color shifts can also be observed in the TIP-2018 dataset [22], but are not very conspicuous in most cases, and can be well tolerated by our layered moiré image model given the offset compensation in the refinement stage.

To investigate the robustness of learning based demoiréing methods against such color shifts, we conduct a cross-domain test with models pre-trained on TIP-2018 [22] on a small dataset MRBI [31], which features diverse device combinations and apparent color shifts within data pairs for joint brightness correction and demoiréing. The reason for the overall low numerical performances in Table 4 ("Cross-domain") mainly lies in that the global color shifts aggravate the difficulty of faithfully recovering the screen contents, given the severely degraded inputs in MRBI [31] compared to reference images. But all learning based demoiréing methods manage to realize obvious quality enhancement for such inputs despite the domain gap. Particularly, our model obviously outperforms other methods in this cross-domain generalization test. And such superiority can also be validated by the qualitative results at the bottom of Figure 13, where our result better suppresses the aberrant moiré color distortions, though the global color shifts cannot be precisely corrected. Such color shift artifacts are far more prominent on the MRBI dataset [31] than TIP-2018 benchmark [22], because the color shifts in MRBI [31] are much more drastic and derived from more diverse capture setup. We can also infer from our results that the color shifts have limited influence on the proposed model to discover and remove moiré patterns. We suppose that better color shift compensation can be realized by color rescaling for further enhancing image restoration results [23].

5.4 Ablation study

In this section, we conduct ablation studies with different model variants to testify to the effectiveness of the proposed module components and constraints. Specifically, for the layer disentanglement design, we change the double-branch module $\mathcal{N}_C$ in the framework to be a single-branch one predicting the clean image $\mathbf{I}_C$ only, which is denoted as **w/o DE**. As for the dual-domain refinement, we first construct a framework variant that skips the frequency domain based moiré residue removal process and attaches the edge attention to the final offset adjustment block, denoted as **w/o FR**, and build another model without the edge attention, notated as **w/o EA**. Additionally, to evaluate the contributions of the proposed new constraints for disentangling the moiré pattern layer, we also train model variants without the self-reconstruction constraint and the low-rank constraint, denoted as **w/o SRC** and **w/o LRC**, respectively. We also look into different choices of the implementation of dual-domain moiré residue removal, which is presented in the supplementary materials.

From the numerical results in Table 5, we can see that the full model achieves the best performance among model variants, and all proposed components contribute to a performance gain, among which the layer disentanglement scheme (w/o DE) most evidently influences the model performance. We assume that its significance in boosting demoiréing performance lies in that the extra branch expands the network capacity, and the proposed constraints for the moiré pattern layer alleviate the ill-posedness in optimization. Besides, we can conclude that the layer disentanglement is also time-efficient, requiring only a little additional inference time, since the two disentanglement branches work in parallel. As for the two proposed constraints, the self-reconstruction constraint is obviously more

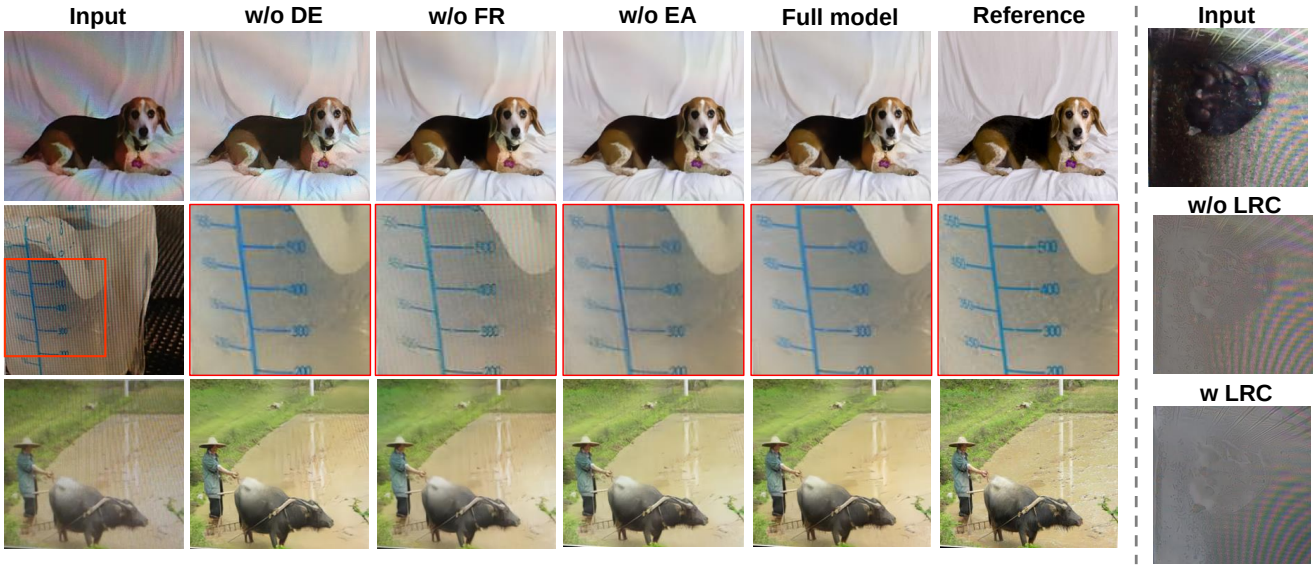


Fig. 14: Visual quality comparison among different model variants. Red boxes mark regions contain slight differences that need a close-up check.

dominant for constraining the disentanglement, because the low-rank constraint only may cause trivial solutions for the moiré pattern layer.

We further investigate the impact of each module component from the perspective of visual results. As shown in the top row of Figure 14, the layer disentanglement is more crucial for suppressing severe color distortions among the proposed components. In the second row, we can observe that without the channel-wise edge attention, the subtle edges of the numbers in the demoiréing result tend to be over-smoothed compared to other ones. Besides, the DCT domain features seem to contribute more to eliminating the periodic high-frequency moiré stripes. In the bottom row, we can see that by aggregating all proposed components, the full model can more thoroughly remove the spatially-varying moiré pattern in the image. But comparing the demoiréing result with the input and reference, we infer that if the recaptured screen image is blurry itself due to unsatisfactory capturing conditions, *i.e.*, the transformation $\mathcal{T}(\cdot)$ to the screen content in Equation 2, the proposed demoiréing framework cannot fully recover the blurred textures, since this is beyond our optimization target of solving for the clean image $\mathbf{I}$ in Section 3.2. In addition, from the right column in Figure 14, we can also see the influence of the proposed low-rank constraint on disentangling a moiré pattern layer with less image textures.

Low-rank constraint on periodic image structures. The proposed low-rank constraint supports the optimization of coarse disentanglement between the image content and the moiré pattern layer, yet there can be special cases where the image itself contains obvious periodic structures, especially in the textured moiré images. In such cases, the low-rank constraint might confront with image texture loss as shown in the bottom row of Figure 15. Similar artifacts also occur in previous non-disentangling methods such as [36], as they tend to smooth the image. But our framework can effectively reduce such unsatisfactory cases by complementing the coarse disentanglement with the refinement stage, as shown in Figure 15, the fabric lines are mostly restored through the edge attention.

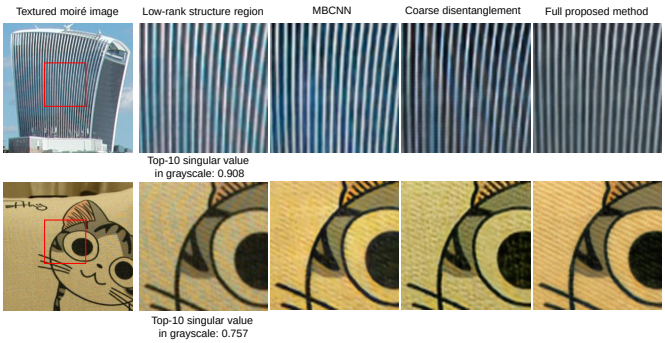


Fig. 15: Two examples of moiré images whose structures contain periodic lines.

Disentanglement for moiré pattern extraction. Though the predicted moiré pattern layer in Figure 8 is not exactly precise, as the low-rank and self-reconstruction constraints cannot fully supervise model learning without the ground truth moiré distortions, the proposed coarse layer disentanglement has shown the capability to describe the moiré pattern's distribution within an image. To further investigate the disentanglement module's potential of explicitly modeling moiré patterns' appearance with more supervision, we additionally capture 3000 image pairs composed of two recaptured screen images $\{\mathbf{I}_M, \mathbf{M}_w\}$ taken with identical capturing conditions. Within such a pair, only the contents $\{\mathbf{I}, \mathbf{I}_w\}$ shown on the screen are different, with $\mathbf{I}_w$ set as a pure white image and $\mathbf{I}$ set as an arbitrary Internet image. Thereby $\mathbf{M}_w$ can be considered to show pure moiré pattern, which has the same contour as the pattern in $\mathbf{I}_M$, as shown in Figure 16. We choose a pure white image as the screen content to highlight the pattern, because moiré patterns are

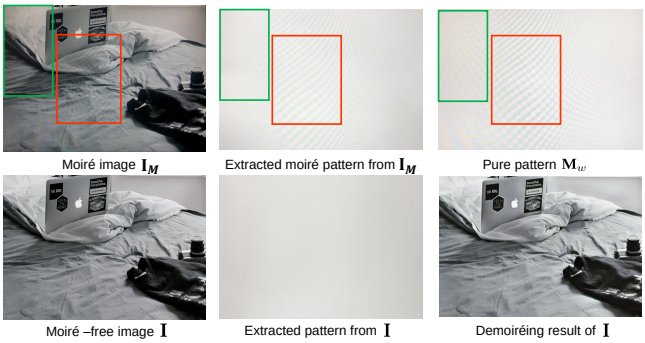


Fig. 16: Top: From left to right are a recaptured screen image $\mathbf{I}_M$, the moiré pattern extracted by the moiré pattern layer disentanglement branch, and the corresponding pure moiré pattern image $\mathbf{M}_w$ paired with $\mathbf{I}_M$, respectively. Bottom: From left to right are the ground truth image $\mathbf{I}$ for $\mathbf{I}_M$, the moiré pattern extraction result, and the demoiréing result for this moiré-free image.

more salient with higher region brightness[4].

To adapt the disentanglement branch to direct pattern extraction, we fine-tune the pre-trained branch for predicting moiré color distortion layer with the newly collected data above. As shown in Figure 16, we can observe that the disentanglement based moiré pattern extraction can successfully depict the prominent pattern structures, as shown in the regions marked by red boxes in Figure 16, which are consistent with the reference pure moiré pattern $\mathbf{M}_w$. Yet different from the pure moiré pattern $\mathbf{M}_w$ that reveals the complete pattern structure, the extracted moiré pattern only focuses on visible color distortions, while tends not to complement the unnoticeable structures according to visible contours. Therefore, it is improper to directly compare the extracted pattern with pure pattern $\mathbf{M}_w$ quantitatively.

The proposed coarse layer disentanglement module is validated to be effective in perceiving the visible moiré patterns from recaptured screen images, which can help it to distinguish the moiré-degraded images from moiré-free ones, *i.e.*, deciding whether to modify the input. We further find that benefiting from such capability, the proposed framework can keep the moiré-free input barely modified, along with a near-blank response in moiré pattern extraction as shown in Figure 16, which demonstrates the reliability of disentanglement based demoiréing.

6 Conclusion

In this paper, we investigate the formation process of moiré patterns from the perspective of signal aliasing, and embody our knowledge on moiré pattern properties in a coarse-to-fine disentangling demoiréing framework for recaptured screen images, which achieves better performance than state-of-the-art methods both quantitatively and qualitatively. The proposed disentanglement based demoiréing shows considerable success on improving model adaptation capability for different data quantities and sources, and we suppose that its advantages comes from the exploitation of moiré pattern related constraints for jointly learning both layers instead of the clean image alone. Thus we believe the coarse disentanglement scheme can be reinforced in future demoiréing works, where more insights on moiré patterns' frequency domain distribution can be proposed to strengthen the learning constraint for the moiré pattern layer. By such means, we would be able to further reduce the dependence on the ground truth images in the training set, and try weakly-supervised learning setups for better generalization across different data sources. Moreover, the coarse-to-fine demoiréing pipeline can also potentially be adapted for similar image restoration tasks such as retinex based low-light image enhancement, considering their commonality in the approximation for layered image model.

4. The statistical relationship between moiré pattern intensity and image brightness is presented in the supplementary materials

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. 2009.
[2] J. Guo and H. Chao. Building dual-domain representations for compression artifacts reduction. In *Proceedings of the European Conference on Computer Vision*, pages 628–644, 2016.
[3] B. He, C. Wang, B. Shi, and L. Duan. FHDe^{2}Net: Full high definition demoireing network. In *Proceedings of the European Conference on Computer Vision*, pages 713–729, 2020.
[4] B. He, C. Wang, B. Shi, and L.-Y. Duan. Mop moiré patterns using MopNet. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2424–2432, 2019.
[5] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7132–7141, 2018.
[6] M. Janner, J. Wu, T. Kulkarni, I. Yildirim, and J. B. Tenenbaum. Self-supervised intrinsic image decomposition. In *Advances In Neural Information Processing Systems*, 2017.
[7] B. Liu, X. Shu, and X. Wu. Demoiréing of camera-captured screen images using deep convolutional neural network. *arXiv preprint, arXiv:1804.03809*, 2018.
[8] F. Liu, J. Yang, and H. Yue. Moiré pattern removal from texture images via low-rank and sparse matrix decomposition. In *IEEE Visual Communications and Image Processing*, pages 1–4, 2015.
[9] L. Liu, J. Liu, S. Yuan, G. G. Slabaugh, A. Leonardis, W. Zhou, and Q. Tian. Wavelet-based dual-branch network for image demoiréing. In *Proceedings of the European Conference on Computer Vision*, pages 86–102, 2020.
[10] L. Liu, S. Yuan, J. Liu, L. Bao, G. G. Slabaugh, and Q. Tian. Self-adaptively learning to demoiré from focused and defocused image pairs. In *Advances in Neural Information Processing Systems*, 2020.
[11] D. Menon and G. Calvagno. Color image demosaicking: An overview. *Signal Processing: Image Communication*, 26(8-9):518–533, 2011.
[12] K. Nishioka, N. Hasegawa, K. Ono, and Y. Tatsuno. Endoscope system provided with low-pass filter for moire removal, Feb. 15 2000. US Patent 6,025,873.
[13] A. V. Oppenheim, A. S. Willsky, and S. Hamid. Signals and systems, second edition. 1997.
[14] I. Pekkucuksen and Y. Altunbasak. Multiscale gradients-based color filter array interpolation. *IEEE Transactions on Image Processing*, 22(1):157–165, 2012.
[15] W. T. Plummer. Anti-aliasing optical system with pyramidal transparent structure, Feb. 5 1991. US Patent 4,989,959.
[16] X. Qin, Z. Wang, Y. Bai, X. Xie, and H. Jia. Ffa-net: Feature fusion attention network for single image dehazing. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence*, pages 11908–11915, 2020.
[17] R. Sasada, M. Yamada, S. Hara, H. Takeo, and K. Shimura. Stationary grid pattern removal using 2d technique for moire-free radiographic image display. In *Proceedings of the SPIE Conference on Medical Imaging*, 2003.
[18] M. Schöberl, W. Schnurrer, A. Oberdörster, S. Fößel, and A. Kaup. Dimensioning of optical birefringent anti-alias filters for digital cameras. In *IEEE International Conference on Image Processing*, pages 4305–4308, 2010.
[19] H. Siddiqui, M. Boutin, and C. A. Bouman. Hardware-friendly descreening. *IEEE Transactions on Image Processing*, 19(3):746–757, 2009.
[20] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[21] B. Sun, S. Li, and J. Sun. Scanned image descreening with image redundancy and adaptive filtering. *IEEE Transactions on Image Processing*, 23(8):3698–3710, 2014.
[22] Y. Sun, Y. Yu, and W. Wang. Moiré photo restoration using multiresolution convolutional neural networks. *IEEE Transactions on Image Processing*, 27(8):4160–4172, 2018.
[23] R. Wan, B. Shi, H. Li, L. Duan, A. Tan, and A. C. Kot. Corrn: Cooperative reflection removal network. *IEEE transactions on Pattern Analysis and Machine Intelligence*, 42(12):2969–2982, 2020.
[24] S. Wang, L. Zhang, and Y. Liang. Nonlocal spectral prior model for low-level vision. In *Proceedings of the Asian Conference on Computer Vision*, pages 231–244, 2012.
[25] Q. Wen, Y. Tan, J. Qin, W. Liu, G. Han, and S. He. Single image reflection removal beyond linearity. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3771–3779, 2019.
[26] N. Yair and T. Michaeli. Multi-scale weighted nuclear norm image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018.
[27] J. Yang, F. Liu, H. Yue, X. Fu, C. Hou, and F. Wu. Textured image demoiréing via signal decomposition and guided filtering. *IEEE Transactions on Image Processing*, 26(7):3528–3541, 2017.
[28] J. Yang, X. Zhang, C. Cai, and K. Li. Demoiréing for screen-shot images with multi-channel layer decomposition. In *IEEE Visual Communications and Image Processing*, pages 1–4, 2017.
[29] S. Yuan, Z. Fu, J. Yang, et al. AIM 2019 challenge on image demoireing: Methods and results. In *IEEE International Conference on Computer Vision Workshops Workshops*, pages 3534–3545, 2019.
[30] S. Yuan, R. Timofte, A. Leonardis, et al. NTIRE 2020 challenge on image demoireing: Methods and results. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1882–1893, 2020.
[31] H. Yue, Y. Mao, L. Liang, H. Xu, C. Hou, and J. Yang. Recaptured screen image demoiréing. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(1):49–60, 2021.
[32] H. Zhang and V. M. Patel. Density-aware single image de-raining using a multi-stream dense network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 695–704, 2018.
[33] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 586–595, 2018.
[34] X. Zhang, Q. Chen, R. Ng, and V. Koltun. Zoom to learn, learn to zoom. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3762–3770.
[35] B. Zheng, Y. Chen, X. Tian, F. Zhou, and X. Liu. Implicit dual-domain convolutional network for robust color image compression artifact reduction. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(11):3982–3994, 2020.
[36] B. Zheng, S. Yuan, G. G. Slabaugh, and A. Leonardis. Image demoireing with learnable bandpass filters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3633–3642, 2020.
[37] B. Zheng, S. Yuan, C. Yan, X. Tian, J. Zhang, Y. Sun, L. Liu, A. Leonardis, and G. Slabaugh. Learning frequency domain priors for image demoireing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Early Access, 2021.

Ce Wang is currently a Ph.D. student in the National Engineering Laboratory of Video Technology (NELVT), School of Electronics Engineering and Computer Science, Peking University (PKU), China. He received his B.E. degree from Nankai University in 2017. His research interests are centered around learning based methods for image processing.

Bin He is currently working at Megvii Technology CO., LTD. He received the Master degree from Peking University in 2021, and his B.S. degree from the Dalian University of Technology in 2018. His current interests include low-level vision and computational photography.

Shengsen Wu is a post-graduate student pursuing a master's degree at Peking University. He received his B.S. degree from Xiamen University in 2019. His current interests include large-scale image retrieval and deep metric learning.

Renjie Wan received his BEng degree from the University of Electronic Science and Technology of China in 2012 and the Ph.D. degree from Nanayang Technological University, Singapore, in 2019. He is currently an Assistant Professor of Hong Kong Baptist University, Hong Kong. He is the outstanding reviewer of ICCV 2019 and the recipient of the Microsoft CRSF Award, VCIP 2020 Best Paper Award, and the Wallenberg-NTU Presidential Postdoctoral Fellowship.

Boxin Shi received the BE degree from the Beijing University of Posts and Telecommunications, the ME degree from Peking University, and the PhD degree from the University of Tokyo, in 2007, 2010, and 2013. He is currently a Boya Young Fellow Assistant Professor and Research Professor at Peking University, where he leads the Camera Intelligence Lab. Before joining PKU, he did research with MIT Media Lab, Singapore University of Technology and Design, Nanyang Technological University, National Institute of Advanced Industrial Science and Technology, from 2013 to 2017. His papers were awarded as Best Paper Runner-Up at ICCP 2015 and selected as Best Papers from ICCV 2015 for IJCV Special Issue. He has served as an editorial board member of IJCV and an area chair of CVPR/ICCV. He is a senior member of IEEE.

Ling-Yu Duan is a Full Professor with the National Engineering Laboratory of Video Technology (NELVT), School of Electronics Engineering and Computer Science, Peking University (PKU), China, and has served as the Associate Director of the Rapid-Rich Object Search Laboratory (ROSE), a joint lab between Nanyang Technological University (NTU), Singapore, and Peking University (PKU), China since 2012. He is also with Peng Cheng Laboratory, Shenzhen, China, since 2019. He received the Ph.D. degree in information technology from The University of Newcastle, Callaghan, Australia, in 2008. His research interests include multimedia indexing, search, and retrieval, mobile visual search, visual feature coding, and video analytics, etc. He has published about 200 research papers. He received the IEEE ICME Best Paper Award in 2019/2020, the IEEE VCIP Best Paper Award in 2019, and EURASIP Journal on Image and Video Processing Best Paper Award in 2015, the Ministry of Education Technology Invention Award (First Prize) in 2016, the National Technology Invention Award (Second Prize) in 2017, China Patent Award for Excellence (2017), the National Information Technology Standardization Technical Committee "Standardization Work Outstanding Person" Award in 2015. He was a Co-Editor of MPEG Compact Descriptor for Visual Search (CDVS) Standard (ISO/IEC 15938-13) and MPEG Compact Descriptor for Video Analytics (CDVA) standard (ISO/IEC 15938-15). Currently he is an Associate Editor of IEEE Transactions on Multimedia, ACM Transactions on Intelligent Systems and Technology and ACM Transactions on Multimedia Computing, Communications, and Applications, and serves as the area chairs of ACM MM and IEEE ICME. He is a member of the MSA Technical Committee in IEEE-CAS Society.

Coarse-to-fine Disentangling Demoiréing Framework for Recaptured Screen Images -Supplementary Material-

Ce Wang[1], Bin He[1], Shengsen Wu[1], Renjie Wan[3], Boxin Shi[1,2], and Ling-Yu Duan[1,2]
[1]Peking University [2]Peng Cheng Laboratory [3]Hong Kong Baptist University

Contents

1. Residue Removal Implementation Choice

In this paper, we propose a DCT based dual-domain refinement backbone featuring the implicit domain transform by the IDCT layer. To further validate our choices of such designs, we compare different variants of the residue removal backbone on the TIP-2018 dataset [5].

The modifications include changing the implicit domain transform to an explicit one [1], [2], with an additional convolutional DCT layer, and using DFT features or spatial features by changing the IDCT layers to dilated convolutional layer or IDFT layer to substitute the DCT features. Particularly, the DFT transform involves complex computation, the feature channels consist of both the real part and the imaginary part of the coefficients transformed by two sets of kernels separately. The numerical comparison results are shown in Table 1. First, we can see that, compared to the explicit transform that fixes more layer weights, the adopted implicit domain transform yields better results, possibly because of the higher degree of freedom in learning priors in the frequency domain. In terms of the dual-domain refinement, the DCT features boosts the demoiréing performance better than its substitutes using FFT and spatial features. We infer that the advantage of DCT and FFT features over pure spatial features originates from their implicit focus across different frequency subbands, and DCT features achieve higher performance possibly because of their better energy compaction without the imaginary part [8].

TABLE 1: Performance comparison among different implementations of feature domain transform on TIP-2018 dataset [5].

	Explicit DCT	FFT	Spatial	Proposed (Implicit DCT)
PSNR	27.468	27.513	27.030	**27.805**
SSIM	0.862	0.861	0.857	**0.863**

2. Capturing Moiré Pattern Extraction Data

The setup for capturing moiré pattern extraction data is shown in Figure 2. In capture, a clean Internet image and a pure white image are sequentially shown in full-screen size on the monitor screen, with the camera phone fixed on a tripod. Thus we can consider the moiré patterns in these two captured screen images have the same structural distribution, since only the screen contents are changed in the capture for a data pair $\{\mathbf{M}, \mathbf{M}_w\}$. For the visibility of moiré patterns in captured image pairs, the distance and viewing angle between the screen and the tripod are adjusted before capturing, to guarantee the clarity of the patterns captured and the whole screen within the viewfinder. The capture is conducted after the parameters are fixed according to current imaging condition, including exposure and focal length. The viewpoint and camera parameters are changed to diversify the data. The viewpoints are slightly moved every five shots within the camera's depth of field, while exposure and focal length are changed every 100 shots.

Moiré pattern intensity and image brightness. We choose a pure white image as the screen content to highlight the pattern contours, considering the saliency of moiré patterns against brighter screen. Such an observation can be validated as a statistical relationship between the moiré pattern's intensity, evaluated by its regional color variation and edge intensity, and the screen content's brightness, as shown in 1.

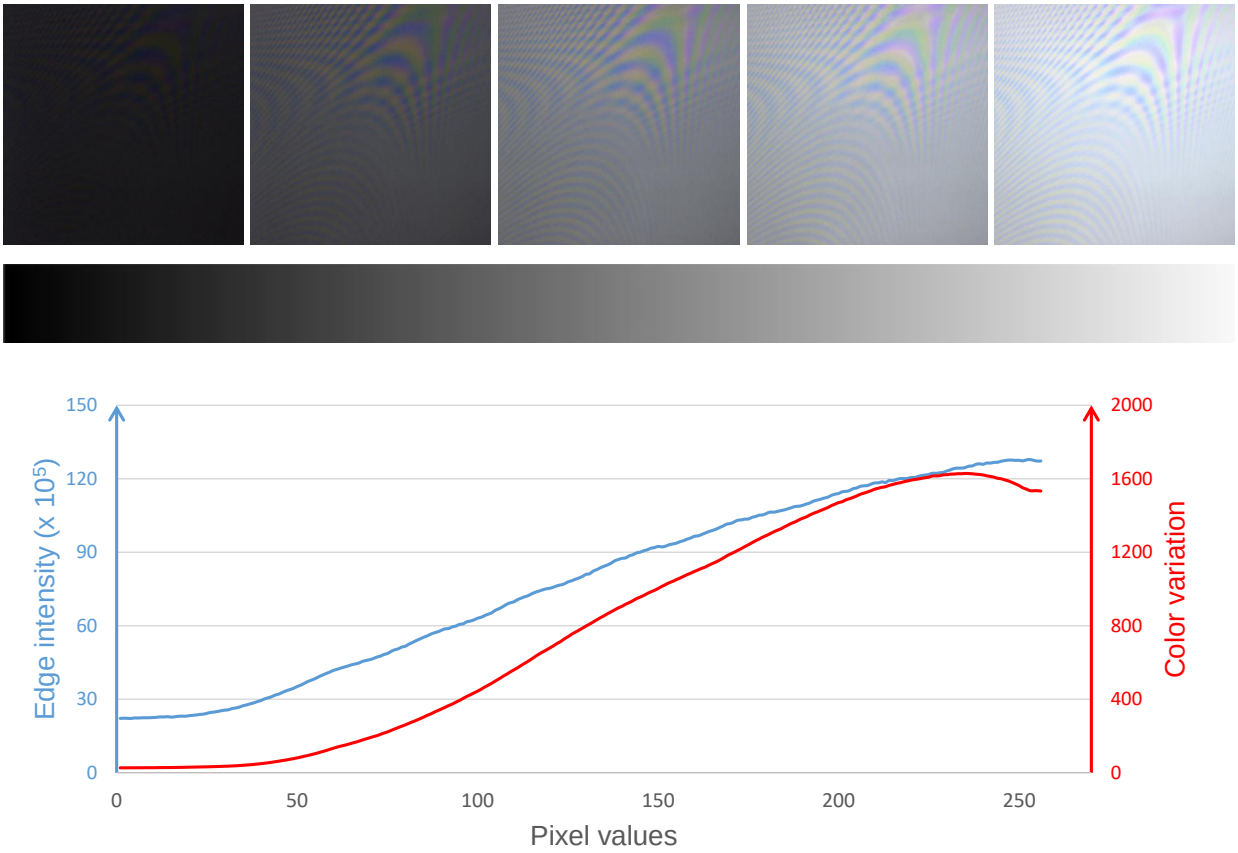
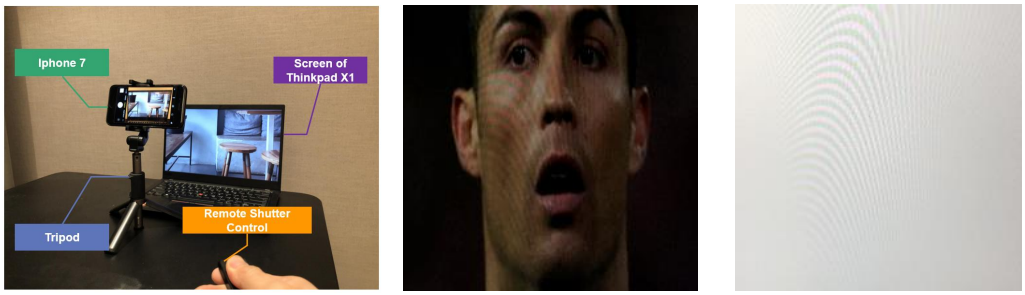


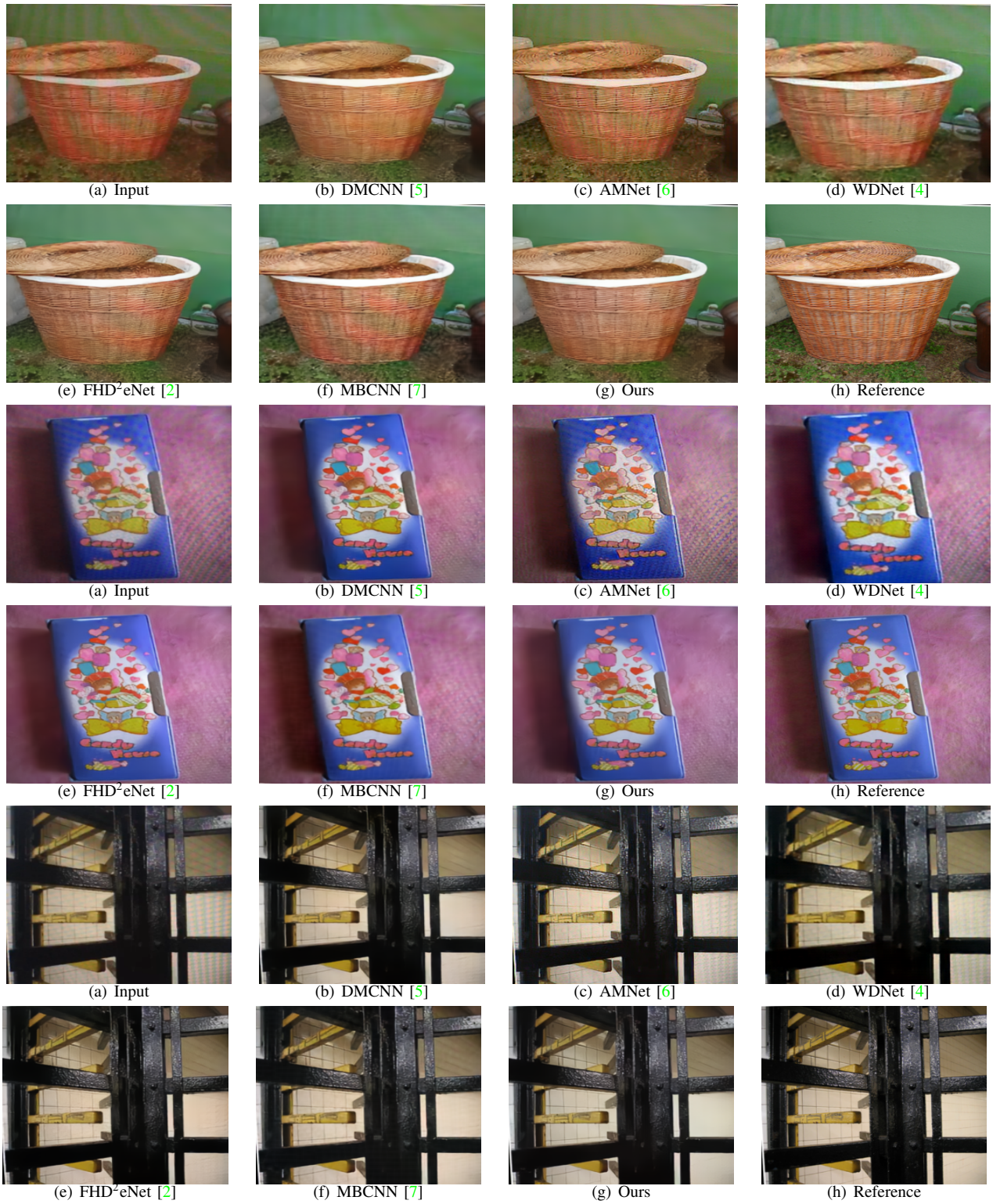
Figure 1: Top: Moiré patterns present different color variation intensity in image regions with different brightness. Bottom: The observed intensity of moiré patterns at the upper part, indicated by color variation and edge intensity, shows a strong positive correlation to the screen contents' brightness.



Capture setup | Moiré image $\mathbf{M}$ | Pure pattern $\mathbf{M}_w$

Figure 2: The data capture setup for moiré pattern extraction (Left), and an additional example of captured image pairs (Middle and Right).

3. Qualitative Results on the TIP-2018 Benchmark



(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

Figure 3: Qualitative results on the TIP-2018 benchmark dataset [5] (Part 1, zoom-in for details).

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

Figure 4: Qualitative results on the TIP-2018 benchmark dataset [5] (Part 2, zoom-in for details).

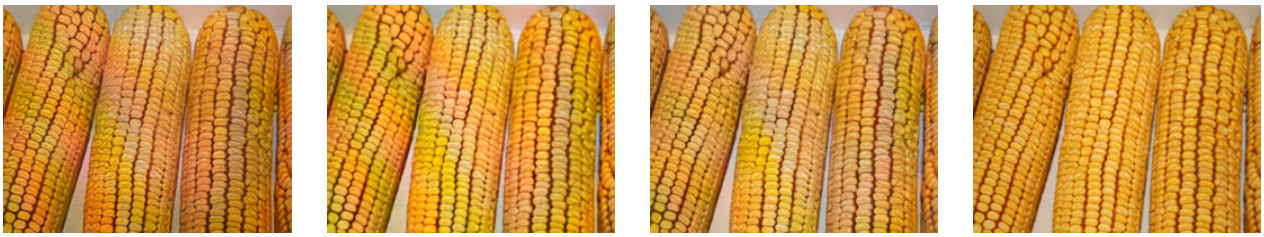
(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

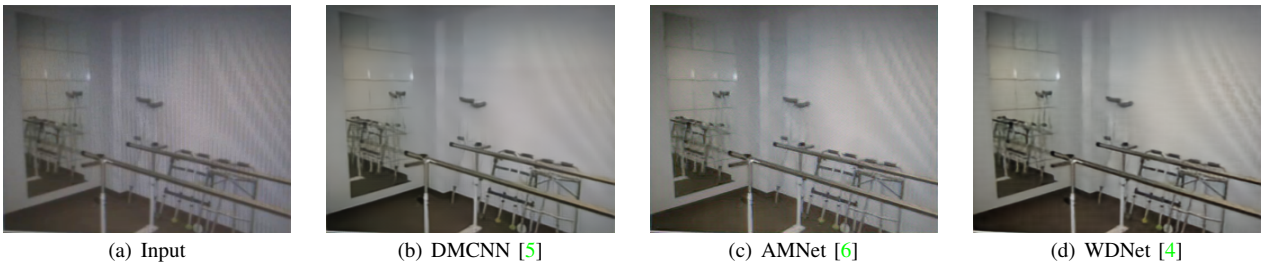
(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

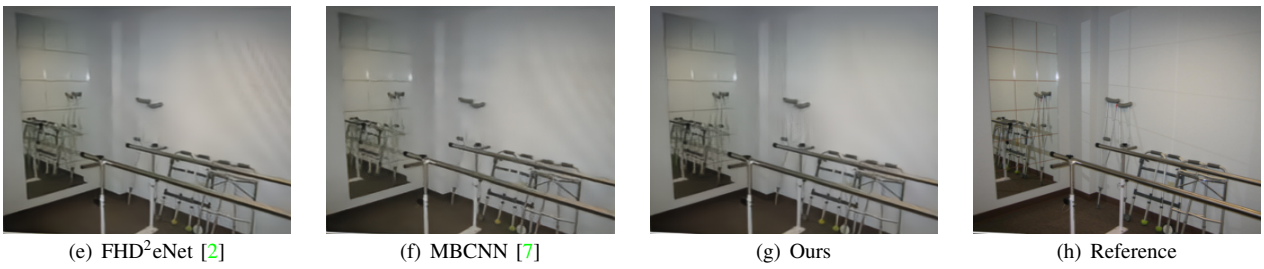


(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

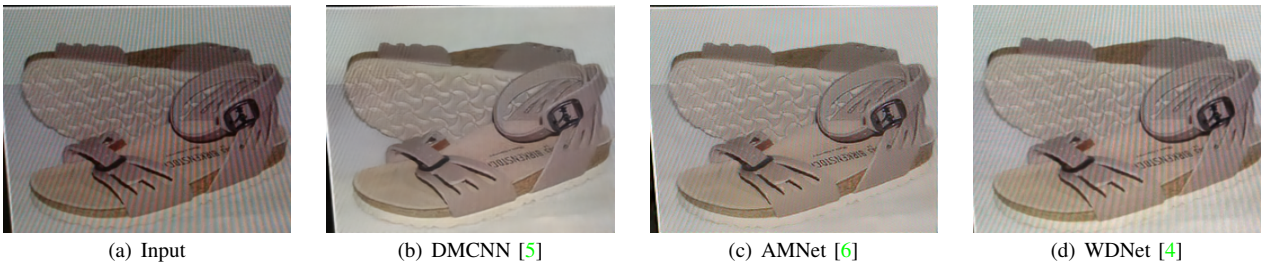
Figure 5: Qualitative results on the TIP-2018 benchmark dataset [5] (Part 3, zoom-in for details).



(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

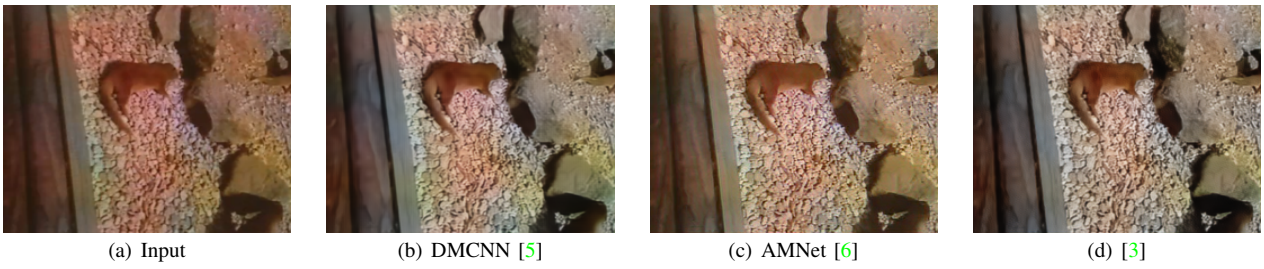


(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

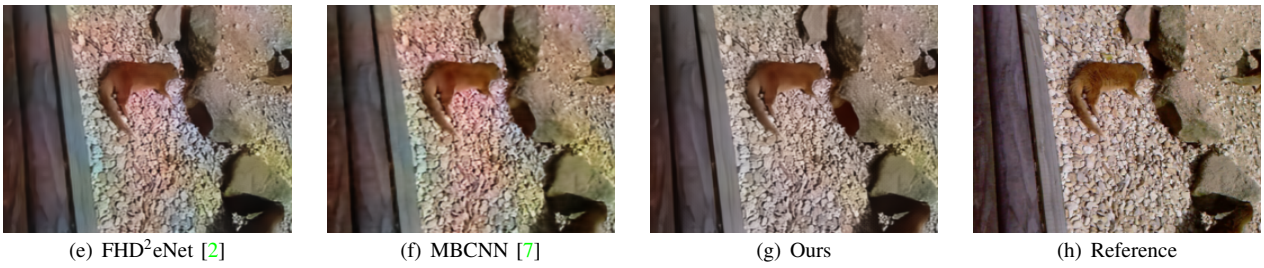


(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

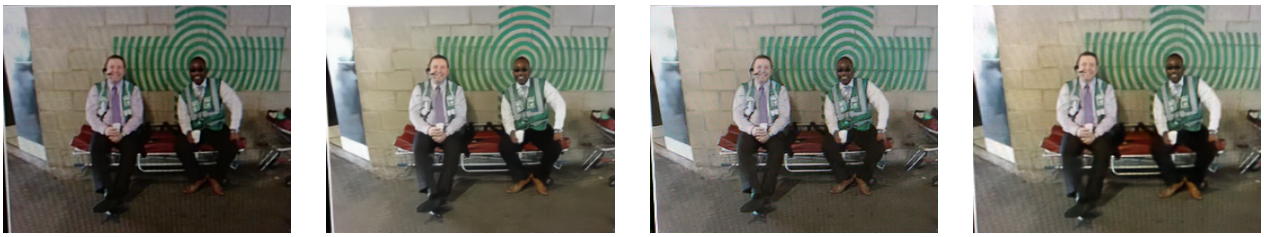


(a) Input (b) DMCNN [5] (c) AMNet [6] (d) [3]

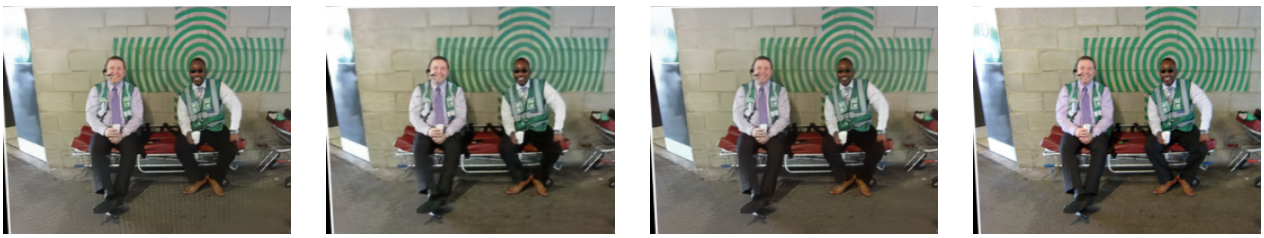


(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

Figure 6: Qualitative results on the TIP-2018 benchmark dataset [5] (Part 4, zoom-in for details).



(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]



(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

(a) Input (b) DMCNN [5] (c) AMNet [6] (d) WDNet [4]

(e) FHD^{2}eNet [2] (f) MBCNN [7] (g) Ours (h) Reference

Figure 7: Qualitative results on the TIP-2018 benchmark dataset [5] (Part 5, zoom-in for details).

4. Detailed Network Architectures

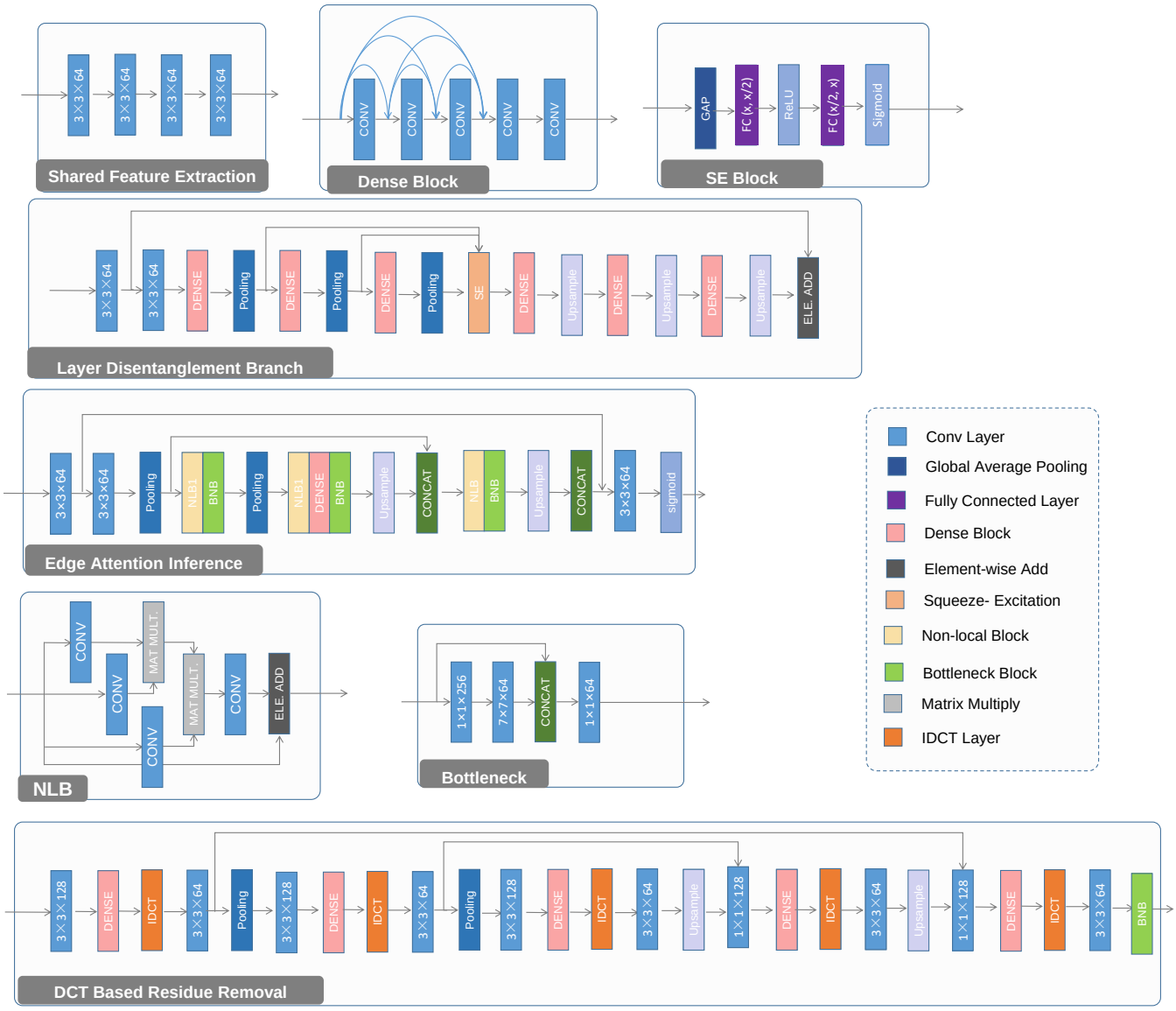


Figure 8: The detailed network architectures, including the functional blocks adopted.

References

[1] J. Guo and H. Chao. Building dual-domain representations for compression artifacts reduction. In *Proceedings of the European Conference on Computer Vision*, pages 628–644, 2016. 2

[2] B. He, C. Wang, B. Shi, and L. Duan. FHDe^{2}Net: Full high definition demoireing network. In *Proceedings of the European Conference on Computer Vision*, pages 713–729, 2020. 2, 4, 5, 6, 7, 8

[3] B. He, C. Wang, B. Shi, and L.-Y. Duan. Mop moiré patterns using MopNet. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2424–2432, 2019. 7

[4] L. Liu, J. Liu, S. Yuan, G. G. Slabaugh, A. Leonardis, W. Zhou, and Q. Tian. Wavelet-based dual-branch network for image demoiréing. In *Proceedings of the European Conference on Computer Vision*, pages 86–102, 2020. 4, 5, 6, 7, 8

[5] Y. Sun, Y. Yu, and W. Wang. Moiré photo restoration using multiresolution convolutional neural networks. *IEEE Transactions on Image Processing*, 27(8):4160–4172, 2018. 2, 4, 5, 6, 7, 8

[6] H. Yue, Y. Mao, L. Liang, H. Xu, C. Hou, and J. Yang. Recaptured screen image demoiréing. *IEEE Transactions on Circuits and Systems for Video Technology*, 31(1):49–60, 2021. 4, 5, 6, 7, 8

[7] B. Zheng, S. Yuan, G. G. Slabaugh, and A. Leonardis. Image demoireing with learnable bandpass filters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3633–3642, 2020. 4, 5, 6, 7, 8

[8] B. Zheng, S. Yuan, C. Yan, X. Tian, J. Zhang, Y. Sun, L. Liu, A. Leonardis, and G. Slabaugh. Learning frequency domain priors for image demoireing. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Early Access, 2021. 2