

# Moiré Photo Restoration Using Multiresolution Convolutional Neural Networks

Yujing Sun, Yizhou Yu, Wenping Wang

**Abstract**—Digital cameras and mobile phones enable us to conveniently record precious moments. While digital image quality is constantly being improved, taking high-quality photos of digital screens still remains challenging because the photos are often contaminated with moiré patterns, a result of the interference between the pixel grids of the camera sensor and the device screen. Moiré patterns can severely damage the visual quality of photos. However, few studies have aimed to solve this problem. In this paper, we introduce a novel multiresolution fully convolutional network for automatically removing moiré patterns from photos. Since a moiré pattern spans over a wide range of frequencies, our proposed network performs a nonlinear multiresolution analysis of the input image before computing how to cancel moiré artefacts within every frequency band. We also create a large-scale benchmark dataset with 100,000+ image pairs for investigating and evaluating moiré pattern removal algorithms. Our network achieves state-of-the-art performance on this dataset in comparison to existing learning architectures for image restoration problems.

**Index Terms**—Moiré pattern, neural network, image restoration

## I. INTRODUCTION

NOWADAYS, digital cameras and mobile phones play a significant role in people’s lives. They enable us to easily record any precious moments that are interesting or meaningful. There exist many occasions when people would like to capture digital screens. Such occasions include taking photos of visual contents on a screen, or shooting scenes involving digital monitors. While image quality is constantly being improved, taking high-quality photos of digital screens still remains challenging. Such photos are often contaminated with moiré patterns (Fig. 4).

A moiré pattern in the photo of a screen is the result of the interference between the pixel grids of the camera sensor and the device screen. It can appear as stripes, ripples, or curves of intensity and colour diversifications superimposed onto the photo. The moiré pattern can vary dramatically due to a slight change in shooting distance or camera orientation. This moiré artefact severely damages the visual quality of the photo. There is a large demand for post-processing techniques capable of removing such artefacts. In this paper, we call images of digital screens taken with digital devices *moiré photos*.

It is particularly challenging to remove moiré patterns in photos, which are mixed with original image signals across a wide range in both spatial and frequency domains. A moiré pattern typically covers an entire image. The colour or thickness of the stripes or ripples in such patterns not only



Fig. 1. Given an image damaged by moiré patterns, our proposed network can remove the moiré artefacts automatically.

changes from image to image, but also is spatially varying within the same image. Thus, a moiré pattern could occupy a high-frequency range in one image region, but a low-frequency range in another region. Due to the complexity of moiré patterns in photos, little research has been dedicated to moiré pattern removal. Conventional image denoising [1], and texture removal techniques [2], [3] are not well suited for this problem because these techniques typically assume noises and textures occupy a higher-frequency band than true image structures.

On the other hand, convolutional neural networks are leading a revolution in computer vision and image processing. After successes in image classification and recognition [4], [5], they have also been proven highly effective in low-level vision and image processing tasks, including image super-resolution [6], [7], demosaicking [8], denoising [9], and restoration [10].

In this paper, we introduce a novel multiresolution fully convolutional neural network for automatically removing moiré patterns from photos. Since a moiré pattern spans over a wide range of frequencies, to make the problem more tractable, our network first converts an input image into multiple feature maps at various different resolutions, which include different levels of details. Each feature map is then fed into a stack of cascaded convolutional layers that maintain the same input and output resolutions. These layers are responsible for the core task of canceling the moiré effect associated with a specific frequency band. The computed components at different resolutions are finally upsampled to the input resolution and fused together as the final output image.

To train and test our multiresolution network, we also create a dataset of 135,000 image pairs, each containing an image contaminated with moiré patterns and its corresponding uncontaminated reference image. The reference images are taken from the ImageNet dataset. The contaminated images

have a wide variety of moiré effects. They are obtained by taking photos of reference images displayed on a computer screen using a mobile phone. To our knowledge, this is the first large-scale dataset for research on moiré pattern removal. The proposed network achieves state-of-the-art performance on this dataset, compared with existing learning architectures for image restoration problems.

We summarise our contributions in this paper as follows.

- 1 We present a novel and highly effective learning architecture for restoring images contaminated with moiré patterns.
- 2 We also create the first large-scale benchmark dataset for moiré pattern removal. This dataset contains 100,000+ image pairs, and will be publicly released for research and evaluation.

## II. BACKGROUND AND RELATED WORK

### A. The Moiré Effect

When two similar, repetitive patterns of lines, circles, or dots overlap with imperfect alignment, a new dynamic pattern appears. This new pattern is called the moiré pattern, which can involve multiple colours. A moiré pattern changes the shape and frequency of its elements when the two original patterns move relative to each other (Fig. 2).

Moiré patterns are large-scale interference patterns. For such interference patterns to occur, the two original patterns must not be completely aligned. Moiré patterns magnify misalignments. The slightest misalignment between the two original patterns could give rise to a large-scale, easily visible moiré pattern. As the degree of misalignment increases, the frequency of the moiré pattern may also increase.

Moiré patterns often occur as an artefact of images generated by digital imaging or computer graphics techniques, such as when scanning a printed halftone picture or rendering a checkerboard pattern that extends toward the horizon [11]. The latter is also a case of aliasing due to undersampling a fine regular pattern.

*a) Moiré Photos:* Photographs of a computer or TV screen taken with a digital camera often exhibit moiré patterns. Examples are shown in Fig. 4. This is because a screen consists of a grid of pixels while the camera sensor is another grid of pixels. When one grid is mapped to another grid, pixels in these two grids do not line up exactly, giving rise to moiré patterns.

Similar to the formation of general moiré patterns, when the relative position between a screen and a digital camera changes, the moiré pattern in the image can change dramatically. It can be 1) of various types: stripes, dots or waves, 2) of various scales, 3) of various levels of intensity, 4) anisotropic or isotropic, and 5) uniform or non-uniform. Removing such moiré patterns with diverse properties is a challenging problem.

The occurrence of moiré patterns in photographs of computer or TV screens does not indicate a defect in the screen but is a result of a practical limitation in display technology. In order to completely eliminate moiré patterns, the dot or stripe pitch on the screen would have to be significantly smaller

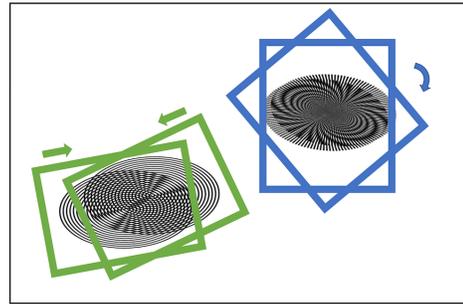


Fig. 2. The mechanism underlying a general moiré pattern. The changing misalignment between two repetitive patterns produces varying moiré patterns.

than the size of a pixel in the camera, which is generally not possible [12].

### B. Related Work

**Moiré Pattern Removal** Several methods have been proposed to remove different types of moiré patterns. Sidorov and Kokaram [13] presented a spectral model to suppress moiré patterns in film-to-video transfer using telecine devices. However, the moiré patterns they deal with are monotonous and monochrome. Thus, their method is unsuitable for eliminating the moiré patterns in our context. Observing that moiré patterns on textures are dissimilar while a texture is locally well-patterned, Liu et al. [14] proposed a low-rank and sparse matrix decomposition method to remove moiré patterns on high-frequency textures as well as on low-frequency structures, the method in [14] is unable to solve our problem. Taking advantage of frequency domain statistics, Sur and Grédiac [15] proposed to remove quasi-periodic noise. Different from our moiré patterns, quasi-periodic noise is simple and regular. Due to the complexity of our moiré patterns, aforementioned methods cannot remove the artefacts well while preserving the original image appearance.

**Image Descreening** In order to print continuous tone images, most electrophotographic printers take advantage of the halftoning techniques, which rely on local dot patterns to approximate continuous tones. Scanned copies of such printed images are commonly corrupted with screen-like high-frequency artefacts (moiré effect), exhibiting low aesthetic quality. Image descreening aims at reconstructing high-quality images from scanned versions of images printed using halftoning (such as scanned books), and has been well studied in the past decades. Various methods have been proposed, such as printer-end algorithms [16], [17], image smoothing techniques [18], learning based methods [19], [20], and advanced filters [21]–[23]. Specialised methods have been proposed to process a specific subset of images, such as paper checks [24]. Shou and Lin [20] descreened images on the basis of a learning based pattern classification process. They found that it is sufficient to consider two classes of moiré patterns to produce satisfactory results. The reason is that halftoning typically involves binary colours, and that the viewing distance and angle during scanning are almost fixed. Such constraints make

moiré patterns in the descreening problem regular, uniform, and local. Therefore, existing image descreening techniques are inadequate to deal with our complex moiré patterns.

**Texture Removal** Since moiré patterns in photos often have high-frequency and repetitive components, texture removal algorithms are a class of relevant techniques. Xu *et al.* [2] introduced relative total variation to describe and identify textures. Karacan *et al.* [25] took advantage of region covariances to separate texture from image structure. Ono *et al.* [26] utilised block-wise low-rank texture characterisation to decompose images into texture and structure components. Cho *et al.* [3] combined the bilateral filter with a “patch shift” texture range kernel to achieve a similar goal. Sun *et al.* [27] took advantage of  $l_0$  norm to retrieve structures from textured images. Ham *et al.* [28] performed texture removal through image filtering with joint static and dynamic guidance. State-of-the-art methods define a variety of local filters to remove high-frequency textures. However, moiré patterns in photos are not merely high-frequency artefacts but span a wide range of frequencies. In addition, moiré patterns also introduce colour distortions, which existing texture removal algorithms would not be able to remove.

**Image Restoration** Image restoration problems aim at removing noises or reconstructing high-frequency details. Recently, learning techniques have been successfully applied to image restoration tasks, including image super-resolution [6], [7], [10], denoising [9], [10], and deblurring [10], [29]. These learning based methods have achieved state-of-the-art performance in image quality improvement. The problem we aim to solve in this paper can be considered as a special image restoration problem as well since it attempts to reconstruct the uncontaminated image by removing moiré artefacts. However, different from the uniformly distributed noises in the denoising task and the missing high-frequency details in the super-resolution task, the moiré patterns in our problem can be anisotropic and non-uniform, and exhibit features across a wide range of frequencies. The models employed in traditional image restoration tasks are not specifically tailored for our problem and can only achieve suboptimal performance. Most recently, Gharbi *et al.* [8] presented a learning-based method to demosaic and denoise images. However, demosaicking is also limited to removing high-frequency artefacts only.

### III. MULTIREOLUTION DEEP CNN FOR MOIRÉ PATTERN REMOVAL

By considering problem complexity, we choose CNNs to remove moiré patterns in photographs due to their recent impressive performance on image restoration tasks. In this section, we present a multiresolution fully convolutional neural network to tackle the problem. It exploits intrinsic correlations between moiré patterns and image components at different levels of a multiresolution pyramid. The training process of our network jointly optimises all parameters to minimize the loss function. As shown in Fig. 1, once trained, our network can automatically remove moiré patterns in contaminated images.

#### A. Network Architecture

Our network architecture is outlined in Fig. 3, which includes multiple parallel branches at different resolutions. The branch at the top processes feature maps at the original resolution of the input image while other branches process coarser and coarser feature maps. The first two convolutional layers in each branch form a group and are responsible for downsampling the feature maps from the immediate higher-level branch by half if there is such a higher-level branch. Therefore the feature maps generated after the first two convolutional layers at all branches can be stacked together to form an upside-down pyramid, where any feature map has half of the resolution of the feature map at the next higher level. Interestingly, in contrast to traditional image pyramids computed using linear filters, our pyramid is computed using nonlinear “filters” (i.e. convolutional kernels + nonlinear activation functions). By converting the input image into multiple feature maps at various different resolutions, we aim to expose different levels of details in the input image.

Inside each branch, the output feature maps from the first two layers are fed into a sequence of cascaded convolutional layers. These convolutional layers maintain the same input and output resolutions, and do not perform any downsampling or pooling operations. They are responsible for the core task of canceling the moiré effect associated with the specific frequency band of that branch. Even with the above multiresolution analysis, this is still a hard task that involves sophisticated nonlinear transforms. Therefore, we place multiple convolutional layers (typically 5) each with  $3 \times 3$  kernels and 64 channels in this sequence.

To assemble the transformed results from all parallel branches together into a complete output image, we still need to increase the resolution of the feature map generated from the cascaded convolutional layers to the original resolution of the input image within each branch except for the first one. In the  $i$ -th branch from the top, we use a set of  $i - 1$  deconvolutional layers to achieve this goal. Each deconvolutional layer doubles the input resolution. There is an extra convolutional layer following the deconvolutional layers within each branch. This extra layer generates a feature map with 3 channels only. This feature map essentially cancels the component of the moiré pattern (in the input image) associated with the frequency band of that branch. At the end, the final 3-channel feature maps from all branches are simply summed together to produce the final output image with the moiré pattern removed.

In our network, whenever there is a need to reduce the resolution of a feature map by half, we use a kernel stride 2 instead of a pooling layer. Each layer is followed by a rectified linear unit (ReLU) and we pad zeros to ensure that the output of each layer is of desired size. The detailed configurations of the first two layers and last layers within all branches are given in Table. I and Table. II, respectively.

*a) Remarks.:* Our deep network is designed on the basis of the key characteristics of moiré patterns, which exhibit features across a wide range of frequencies. A moiré pattern is typically spatially varying and spreads over an entire image. If a network deals with fine-scale features only, low-frequency

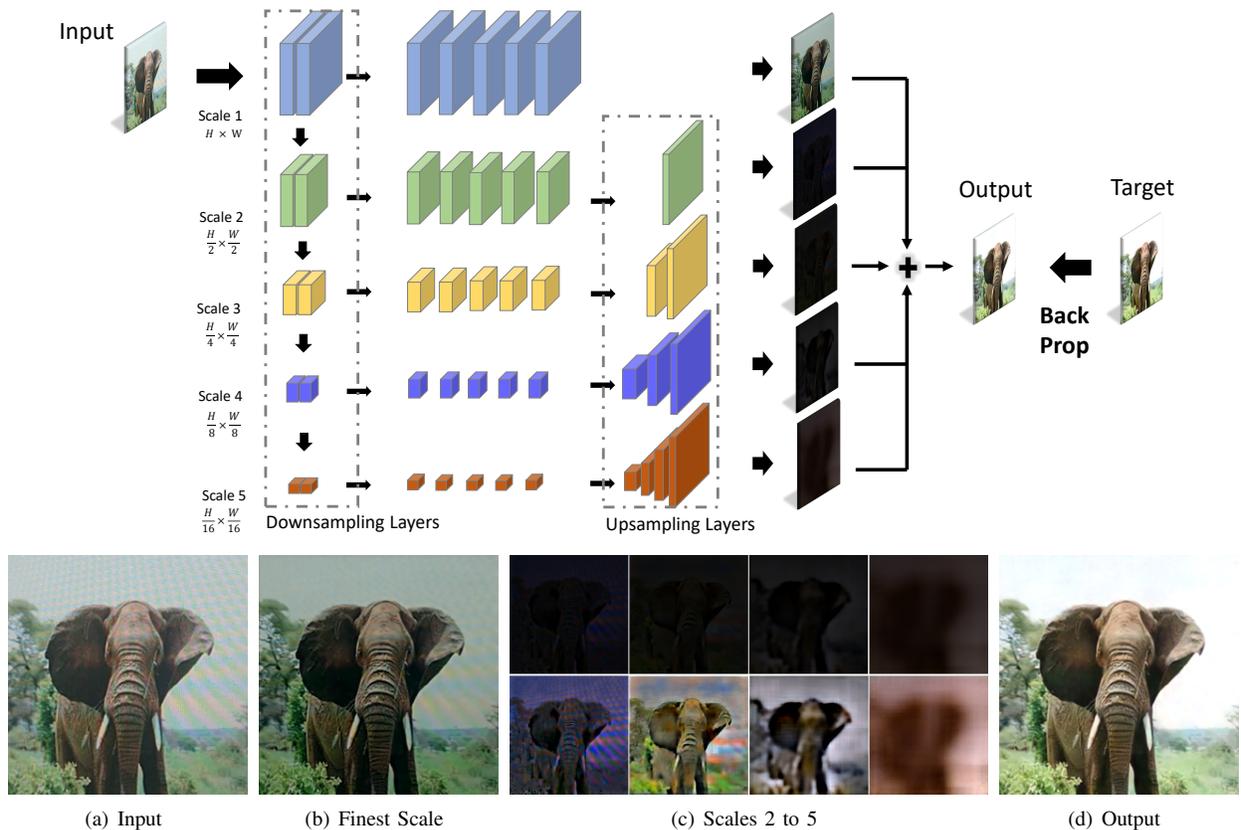


Fig. 3. The architecture of our multiresolution fully convolutional network. The top row in (c) shows intermediate images produced from the second to fifth network branch, and the bottom row shows the same images with amplified intensity.

TABLE I  
DOWNSAMPLING LAYERS

Scale	Kernel	Stride	Channels
1	3x3	1x1	32
1	3x3	1x1	32
2	3x3	2x2	32
2	3x3	1x1	64
3	3x3	2x2	64
3	3x3	1x1	64
4	3x3	2x2	64
4	3x3	1x1	64
5	3x3	2x2	64
5	3x3	1x1	64

TABLE II  
UPSAMPLING LAYERS

Scale	Type	Kernel	Stride	Channels
1	conv	3x3	1x1	3
2	deconv	4x4	2x2	32
	conv	3x3	1x1	3
3	deconv	4x4	2x2	64
	deconv	4x4	2x2	32
	conv	3x3	1x1	3
4	deconv	4x4	2x2	64
	deconv	4x4	2x2	32
	deconv	4x4	2x2	32
	conv	3x3	1x1	3
5	deconv	4x4	2x2	64
	deconv	4x4	2x2	32
	deconv	4x4	2x2	32
	deconv	4x4	2x2	32
	conv	3x3	1x1	3

components of the moiré pattern cannot be removed; if it deals with coarse-scale features only, high-frequency features of the moiré pattern cannot be removed. For these reasons, we perform a multiresolution analysis of the input image and remove the component of the moiré pattern within every frequency band separately.

In Fig. 3, we illustrate how our network removes a moiré pattern from a contaminated image. The network branch for the original resolution (the finest scale) plays a dominant role because pixel colours in the final output image mostly come from this branch. We can see that moiré artefacts have not been completely removed in the 3-channel feature map produced from the last layer of the top branch (Fig. 3(b))

though such artefacts have become much weaker than those in the original input (Fig. 3(a)). Network branches for other coarser resolutions play a supporting role. The last layer of each coarser-resolution branch produces an image that aims to cancel the remaining moiré pattern (in the image produced from the last layer of the top branch) which falls into its frequency band (Fig. 3(c)). When images from all the branches are summed together, the remaining artefacts in the image from the top branch can be successfully eliminated (Fig. 3(d)).

### B. Network Training

We train our deep network using a dataset of images,  $D = \{(I_i, O_i)\}$ , where  $I_i$  is an image contaminated with a moiré pattern and  $O_i$  is its corresponding ground-truth uncontaminated image. The training process solves for weights  $\mathbf{w}$  and biases  $\mathbf{b}$  in our network via minimising the following  $l_2$  loss defined on image patches of size  $p \times p$  from the training set  $D$  in an end-to-end fashion:

$$L(\{\mathbf{w}, \mathbf{b}\}) = \frac{1}{N} \sum_{i=1}^N \|S_i - T_i\|^2, \quad (1)$$

where  $N$  is the total number of image patch pairs and  $(S_i, T_i)$  is a pair of patches.

## IV. DATASET

We create a benchmark of 135,000 image pairs, each containing an image contaminated with a moiré pattern and

its corresponding uncontaminated reference image. The contaminated images have a wide variety of moiré effects (Fig. 4). The uncontaminated reference images in our benchmark come from the 100,000 validation images and 50,000 testing images of the ImageNet ISVRC 2012 dataset. Of the 135,000 pairs of images, 90% are used as the training set and 10% are used for validation and testing. The pipeline to collect this data is shown in Fig. 5, which mainly consists of two steps: image capture and alignment.

*a) Image Capture:* Each reference image is enhanced with a black border and displayed at the centre of a computer screen (Fig. 5(a)). The reason to use black for the border is that we observe dark colours are least affected by the moiré effect. To increase the number of corner points that can be used during image alignment, we further extrude a black block from every edge of the black border. We then fill the rest of the screen outside the black border (and blocks) with pure white, which enables us to easily detect the black border in the captured images. We capture displayed images using a mobile phone (Fig. 5(d)). During image acquisition, we randomly change the distance and angle between the mobile phone and the computer screen. Note that we require the black image borders to be always captured.

Detailed information of the phone models and the monitor screens is shown in Table III and Table IV, respectively. For each combination of phone model and screen, we collected 15,000 pairs of images. Thus, we collected  $15,000 \times 9 = 135,000$  image pairs in total. Using different phone models as our capture devices ensures that moiré patterns are captured across different optical sensors while the diversity of display screens exhibits the difference in screen resolution.

*b) Image Alignment:* The prepared reference images and their corresponding captured images contaminated with moiré patterns have different resolutions and perspective distortions. To train our deep network in an end-to-end manner, we need to register them.

In practice, we rely on the corners along the black image border to accomplish image alignment. Since we use a flat computer screen, the four corners of a captured image (excluding the blocks extruded from the border) lie on a

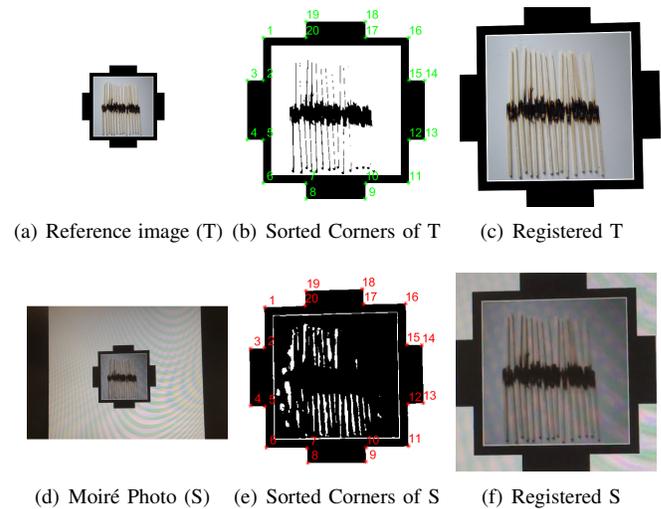


Fig. 5. Image Acquisition.

plane. So do the four corners of the prepared reference image. Therefore, corresponding points in both the captured image and reference image are associated via a homography, which can be represented with a  $3 \times 3$  projective matrix with 8 degrees of freedom. The four black blocks we attached to the image border increase the number of non-collinear corresponding points from 4 to 20, which can improve the registration precision. We use these 20 corners to compute the projective matrix and further align every pair of images.

To detect the corners, we convert the images into binary images and search for corners along the outermost boundary of the black image border. Traditional corner detection methods, such as the Harris corner detector [30], can faithfully detect all corners in a target image (Fig. 6(a)). However, because of the presence of moiré artefacts, they fail to robustly find the 20 corresponding corners in the source image (Fig. 6(b)), where certain edge pixels can be falsely detected as corners.

To eliminate such false corners, we check the ratio between the numbers of black pixels and white pixels in a square neighbourhood around each detected corner. Since each corner forms a right angle, ideally, the ratio between the numbers of



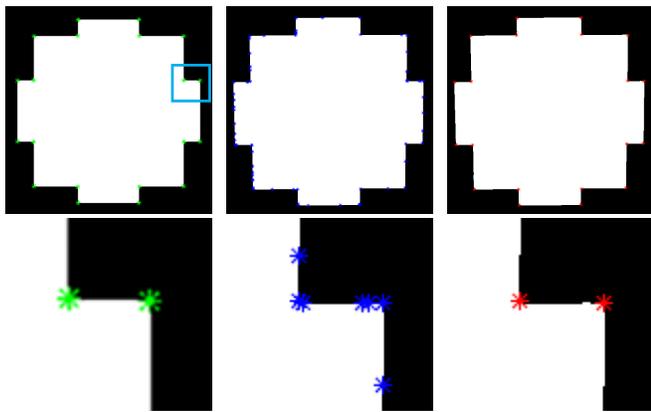
Fig. 4. Examples of image pairs from our dataset. From left to right: images are contaminated by stripe, dot and curved moiré patterns respectively.

TABLE III  
PHONE MODEL SPECIFICATIONS

Manufacturer	Model	Camera
APPLE	iPhone 6	8MP
SAMSUNG	Galaxy S7 Edge	12MP
SONY	Xperia Z5 Premium Dual	23MP

TABLE IV  
DISPLAY SCREEN SPECIFICATIONS

Manufacturer	Model	Resolution	Size (inch)
APPLE	Macbook Pro Retina	$2560 \times 1600$	13.3"
DELL	U2410 LCD	$1920 \times 1200$	24"
DELL	SE198WFP LCD	$1280 \times 800$	19"



(a) Detected corners of T (b) Detected corners of S (c) Cleaned corners of S

Fig. 6. Corner Detection and Clearance.

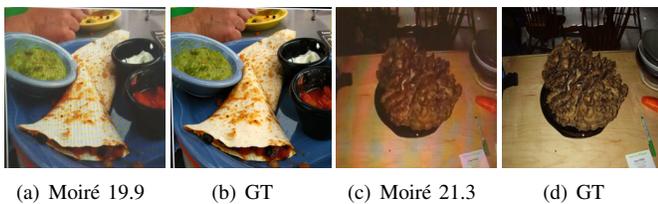


Fig. 7. PSNR cannot fully reflect the degree of moiré patterns. An image corrupted by visually more severe moiré patterns can have higher PSNR.

black and white pixels should be either 3 or  $1/3$ . According to this observation, we filter out false corners, where the ratio between the numbers of black and white pixels in a square neighbourhood is clearly different from 3 or  $1/3$ . In practice, we set the neighbourhood size to  $11 \times 11$ . To remove duplicate corners, we set a minimum distance between two distinct corners. When the pairwise distances among two or more detected corners fall below this threshold, we only keep one of them. As shown in Fig. 6(c), these twenty corners can be successfully detected.

Finally, with the computed projective matrix, we can align every image pair. The registration results are demonstrated in Fig. 5(c) and 5(f).

*c) Automatic Verification:* To automatically verify whether a registration result is correct or not, we measure the PSNR of the registered image pair and use a threshold  $\eta$  to screen the PSNR value. In our experiments, we set  $\eta = 12$ . We have found that even images with the most severe moiré artefacts achieve PSNR values higher than 12dB while false registrations produce PSNR values lower than 10dB. The quality distribution of moiré photos in our dataset is shown in Fig. 8.

However, note that PSNR cannot fully reflect the severity of the moiré effect. As shown in Fig. 7, an image corrupted by a visually more severe moiré pattern actually achieves a higher PSNR. This is perhaps because the colour bands in a moiré pattern do not significantly affect PSNR even though they are visually disturbing and easily noticeable.

*d) Setup:* During image acquisition, images are displayed on the screen consecutively. Each reference image stays

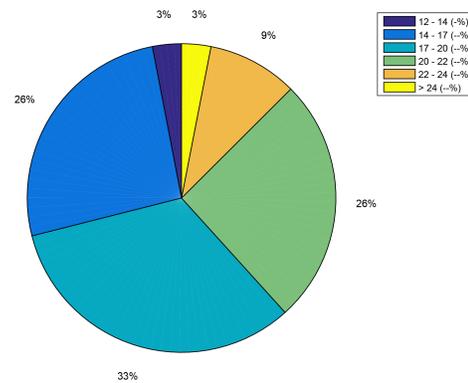


Fig. 8. The quality distribution of moiré photos in the entire dataset. The quality of a moiré photo with respect to its corresponding reference image is measured using PSNR (dB).

on the screen for 0.3 seconds. We use a mobile phone to record a video of the consecutively displayed images. Frames from the captured video are then retrieved as images contaminated with moiré patterns.

## V. MODEL UNDERSTANDING AND IMPLEMENTATION

### A. Insights Behind Our Network Design

Moiré patterns span a wide range in both spatial and frequency domains. Therefore, we conceive a multi-resolution architecture, which has convolutional layers with multi-scale receptive fields, to tackle this problem. At the beginning, we experimented with U-Net [31] with skip connections. Skip connections have been proven to be effective in high-level vision tasks, such as image recognition and semantic segmentation. However, when tackling low-level vision problems, including super-resolution, denoising and deblurring, many approaches can produce state-of-the-art results without skip connections, such as VDSR, DnCNN and PyramidCNN. In high-level vision problems, the information from high-resolution layers close to the input image is useful for the additional clues they introduce. Different from other tasks making use of networks with skip connections, moiré photos and their corresponding ground-truth images can differ dramatically, and thus, skip connections are not powerful enough to model such differences. In addition, the layer closer to the input image in a skip connection contains serious moiré artefacts, as shown in the top row of Fig. 10, while the feature maps produced by the deeper layer are relatively moiré-free. As a result, directly using high-frequency details from a layer closer to the input image would likely introduce artefacts in the final result.

PyramidCNN [29] also adopts a multi-resolution architecture for deblurring. In their architecture, an input image is first downsampled to  $k$  resolutions linearly and then network branches for different resolutions are trained simultaneously. For the task of deblurring, coarser level output guides the training process of finer level network branches. But for moiré pattern removal, the output from coarser levels is not completely free of moiré artefacts, which tend to make finer levels maintain such artefacts.

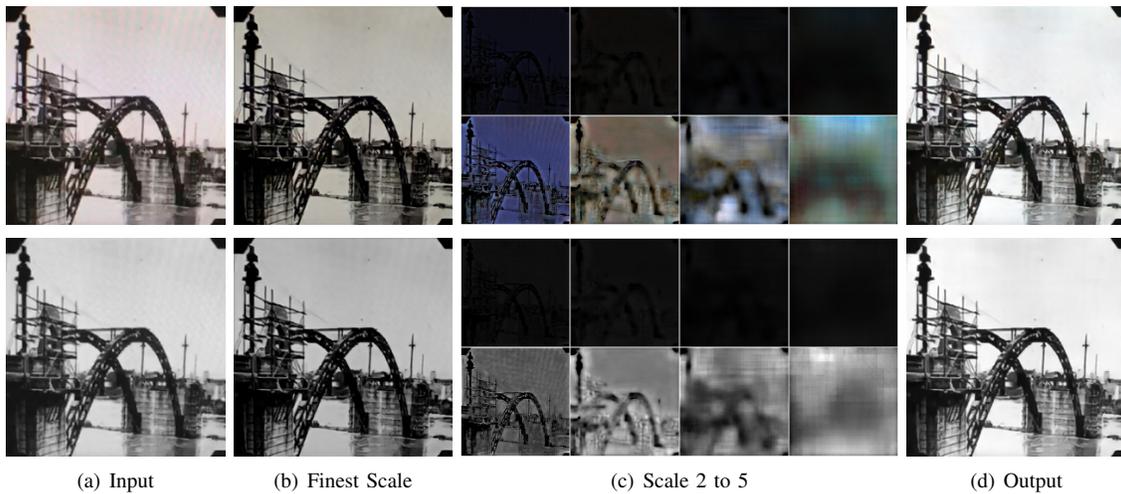


Fig. 9. Visualisation of the 3-channel feature maps produced by different branches on a “grayscale-like” RGB image and its corresponding pure grayscale image. For each input, the top row in (c) shows the intermediate images produced from the second to the fifth network branch, and the bottom row shows the same images with amplified intensity.



Fig. 10. Visualisation of U-Net feature maps. (Top) Feature maps produced by a layer A closer to the input image. (Bottom) Feature maps produced by a deeper layer B. Layer A is skipped connected to layer B.

To achieve better performance, we embed a multi-resolution pyramid in our network architecture. In contrast to traditional image pyramids built with linear filtering, the image pyramid in our architecture is actually built with nonlinear filtering because nonlinear activation always follows each convolutional layer. The nonlinearity in our pyramid allows the network to perform more effectively during downsampling. More importantly, in our network, each resolution is associated with a network branch with six stacked convolutional layers maintaining the same resolution. Such network branches are capable of performing sophisticated nonlinear transformations (such as removing moiré artefacts within a specific frequency band), and are more powerful than skip connections in U-Net.

### B. A Detailed Study on Our Proposed Model

To show the advantage of the proposed model, we attempt to test different variants. Model specifications are given as follows:

- V\_Concate (27.12dB): replacing the sum operation with concatenation. To be specific, we concatenate the 32 feature maps from each scale, and append two convolutional

layers after the concatenated feature maps. Each of these convolutional layers has 32 channels and  $3 \times 3$  kernels.

- V\_Skip (26.36dB): in each scale, skip connecting the second downsampling layer to the last convolutional layer before the upsampling layers.
- V\_C32 (25.52dB): replacing all the 64-channel convolution filters with 32 channel convolutional filters.
- V\_B123 (25.28dB): using branch 1, 2 and 3 only.
- V\_B135 (26.04dB): using branch 1, 3 and 5 only.
- V\_B15 (25.52dB): using branch 1 and 5 only.

We will demonstrate later that although V\_Concate achieves a higher PSNR score on the test data, it produces worse visual results than our proposed network. Adding skip connections cannot further improve the performance of the proposed model while the other variants degrade the performance.

### C. Grayscale Moiré Artefacts

To verify that our model can remove moiré patterns rather than the unnatural colours, we convert the RGB dataset to a grayscale one and retrain the network. The average PSNR, SSIM and FSIM on the grayscale testing set are 27.26, 0.852, and 0.910, respectively, indicating that our model is able to deal with moiré patterns regardless of the colour information. Intermediate images produced from different branches on a test RGB image that is close to a grayscale one as well as those produced on its corresponding grayscale image are demonstrated in Fig. 9.

### D. Implementation

We have fully implemented our proposed deep multiresolution network using Caffe on an NVIDIA Geforce 1080 GPU. The entire training process takes 3 days on average. We use a mini-batch size of 8, start with learning rate 0.0001, set the weight decay to 0.00001, and minimize the loss function using Adam [34]. We have found that the training process could not converge properly with a higher learning rate. As

TABLE V

A QUANTITATIVE COMPARISON AMONG PARTICIPATING METHODS ON OUR TEST SET WITH DIFFERENT METRICS. OUR METHOD CLEARLY OUTPERFORMS THE OTHER METHODS.

	Corrected Input	RTV [2]	SDF [28]	IRCNN [10]	DnCNN [9]	VDSR [7]	PyramidCNN [29]	U-Net [31]	V_Concate	Our method
PSNR Mean (dB)	20.30	20.67	20.88	21.01	24.54	24.68	25.39	26.49	27.12	26.77
PSNR Gain (dB)	-	0.37	0.58	0.71	4.24	4.38	5.09	6.19	7.09	6.47
Ave Error ( $\times 10^{-3}$ )	34	31	30	28.32	5.82	5.74	4.83	3.81	3.36	3.62
SSIM [32]	0.738	-	-	-	0.834	0.837	0.859	0.864	0.878	0.871
FSIM [33]	0.869	-	-	-	0.901	0.902	0.909	0.912	0.922	0.914

the training process proceeds, we reduce the learning rate by a factor of 10 when the loss on a validation set stops decreasing. In all the experiments in this paper, we set the patch size  $p \times p$  to  $256 \times 256$ . The network weights are randomly initialised using a Gaussian with a zero mean and a standard deviation equal to 0.01. The bias in each neuron is initialised to 0.

## VI. COMPARISON AND DISCUSSION

In this section, we experimentally analyse our method’s capability in improving image quality and removing moiré artefacts. Since we are not aware of any existing methods that solve exactly the same problem, we compare our method against state-of-the-art methods in related image restoration problems, including image denoising, deblurring, super-resolution and texture removal. We choose VDSR [7] as a representative from image super-resolution algorithms, DnCNN [9] and IRCNN [10] from the latest image denoising methods, and RTV [2] and SDF [28] among texture removal techniques. For that a subset of the moiré photos in our dataset has a certain degree of blurriness and that deblurring techniques can reconstruct high-frequency details, we also add two latest learning based image deblurring techniques, multi-scale pyramidCNN [29] and IRCNN [10], for comparison. Moreover, since we adopt a hierarchical network architecture, we also compare our network with U-Net [31], an effective neural network for image segmentation.

To perform a fair comparison, we tune the parameters of the methods we compare against so that they reach the optimal performance on our dataset. When a method only has a small number of tuneable parameters, we tune those parameters to make the method achieve the lowest average error on our test set. When a method has a large number of parameters, such as learning based methods, we retrain the model in the method using our training set.

Even though descreening methods aim at removing a different and simpler moiré effect that occurs in scanned copies of printed documents and images, they are certainly relevant. Since such methods are relatively mature and have been integrated into commercial software, we choose to compare with the descreening function in Photoshop.

### A. Quantitative Comparison

In Fig. 11 and Table V, we demonstrate the quantitative performance of different methods on our test set. Since the contaminated image and the reference image within the same pair have different average intensity levels due to multiple

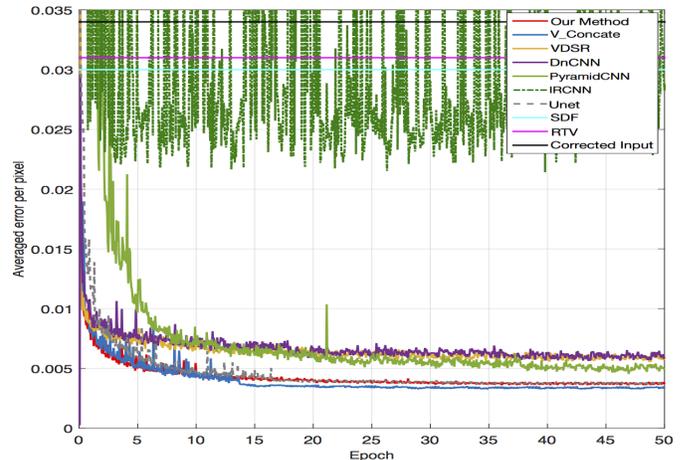


Fig. 11. Average pixel-wise MSE error of various methods vs. the number of epochs.

reasons, including the brightness of the computer screen and the intensity response curve of the camera during image acquisition, that are mostly irrelevant to the moiré effect, we decided to factor out the differences in average intensity by adjusting the average intensity of the contaminated image to be the same as that of the reference image (Corrected Input). As shown, our method and the variant of our model, V\_Concate, outperform all other methods participating in the comparison on all performance measures, including PSNR, SSIM [32] and FSIM [33]. As the parameters for descreening in Photoshop have to be adjusted manually for each image, we cannot show the average performance on the entire test set. However, we will qualitatively compare it with our method in the next section.

Effective as a super-resolution method, VDSR [7] delivers a reasonable performance but is unable to fully handle the complex moiré effect. Using a configuration with a large receptive field, the denoising network (DnCNN) in [9] has a similar performance as VDSR [7]. Both VDSR and DnCNN adopt a flat CNN architecture that maintains the same resolution across all layers. Nonetheless, both of them have been clearly outperformed by our multiresolution network.

By defining a denoising prior with dilated convolutions, IRCNN [10] outperforms state-of-the-art methods in pixel-wise image restoration tasks. However, it performs poorly on our dataset and its training process can hardly converge on our training set. After modifying IRCNN by interleaving ordinary

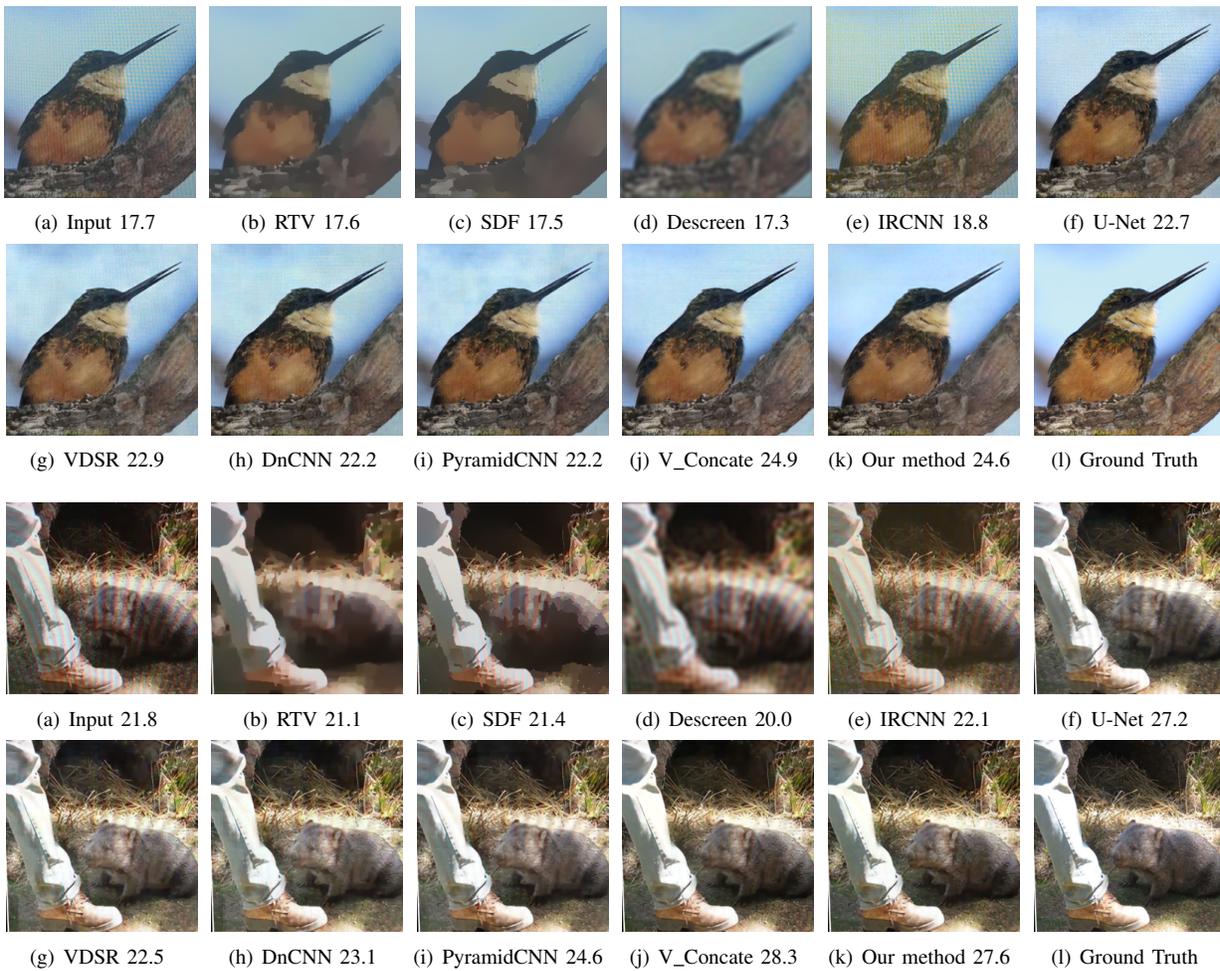


Fig. 12. Comparison between our multiresolution deep network and other state-of-the-art methods for image restoration, including Photoshop Descreen, IRCNN [10], U-Net [31], VDSR [7], DnCNN [9], pyramidCNN [29], RTV [2] and SDF [28].

convolutions and dilated convolutions, we obtain a revised model called IRCNN-IL. The convergence issue is resolved in the revised model but its performance is still not satisfactory. The PSNR, SSIM and FSIM achieved by IRCNN-IL are 21.55, 0.744, and 0.870, respectively. In theory, the noise IRCNN aims to deal with is completely different from the moiré patterns we attempt to remove. A noisy image is commonly modelled as the result of an additive process, which adds noise to the original signal, but a moiré pattern is a phenomenon caused by light interference, which is a different and much more complicated process. Dilated kernels can remove additive noises but might be insufficient to remove complex moiré patterns. Due to the different underlying mechanisms of image noises and moiré patterns, one cannot be certain that IRCNN is effective for restoring moiré photos.

Nah *et al.* [29] deblurs images bottom up using a multiresolution Gaussian pyramid. It first deblurs an image in  $1/2^i$  resolution, then in  $1/2^{i-1}$  resolution and finally in the full resolution. The multiresolution architecture helps to produce acceptable results. However, unlike our multiresolution pyramid generated from trainable nonlinear filters (convolutional kernels), their pyramid is generated using the fixed Gaussian filter, which is linear. As shown in Fig. 11 and Table V, our

network architecture delivers clearly better performance.

Among all the methods, U-Net [31] achieves a numerical performance closest to our method. However, we found that even though U-Net produces good statistics, it delivers relatively poor visual results, which will be demonstrated in visual comparisons. Likewise, V\_Concate produces the highest score on all metrics but its ability in visually removing moiré patterns is less than the original model.

Texture removal techniques, RTV [2] and SDF [28], are useful in preserving important image structures while eliminating small repetitive textural details. But image features at a similar scale of texture elements would be removed as well. In our context, these techniques are used for removing moiré patterns, and they give a poor performance on this task. The difficulty in setting an appropriate texture kernel size could be the main reason because a large smoothing and texture kernel would over-smooth the image while a small kernel would not be able to remove low-frequency large-scale moiré artefacts.

### B. Visual Comparisons

We visually compare results from our method against those from other state-of-the-art methods in Fig. 12. Additional

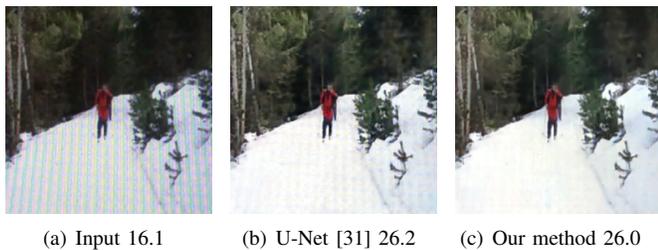


Fig. 13. Another example in which U-Net [31] produces a higher PSNR score but a worse moiré removal effect.

visual comparisons can be found in the supplemental materials. Note that the input images are all from the test set. From these comparisons, we have the following observations. RTV [2] and SDF [28] remove small-scale texture features which typically have higher frequencies than moiré patterns. Descreening in Photoshop over-smoothes the input image. Among deep learning based methods, IRCNN [10] is unable to remove moiré patterns at all even though its network has been re-trained using our training set. Meanwhile, VDSR [7], PyramidCNN [29], and DnCNN [9] have a better performance. However, colour distortion is still noticeable in their results.

Except for our methods, U-Net [31] achieves the highest scores of all quality measures. But more moiré artefacts remain in its results than in the results of VDSR [7] and DnCNN [9]. As we have stated earlier, even though a quality measure, such as PSNR, can measure the overall image quality, it cannot precisely measure the effectiveness in moiré pattern removal. We show an example in Fig. 13 and the supplemental materials that U-Net [31] produces higher PSNRs but worse visual results. Our method has the most powerful network architecture and produces output images closest to the ground-truth reference images.

Additional visual results from our method are shown in Fig. 19, where the input images exhibit a variety of different moiré patterns.

### C. The Number of Variables

As shown in Table. VI, the number of variables in our method is in the same order as U-Net and PyramidCNN while our proposed network outperforms both of them qualitatively and quantitatively. Variants of our model, V\_B15 and V\_C32, have a similar number of parameters as VDSR and DnCNN, however produce higher PSNR scores.

TABLE VI

THE NUMBER OF VARIABLES IN LEARNING BASED APPROACHES ( $\times 10^5$ ).

	V_B123	V_B15	V_C32	V_Concate	Our method
# var	9.28	7.42	4.11	16.14	15.44
	IRCNN-IL	VDSR	DnCNN	PyramidCNN	U-Net
# var	3.35	6.67	7.04	14.15	24.62

### D. User Study

Due to the limitation of image metrics in measuring moiré artefacts, we have also conducted a user study to compare

different methods, which includes 20 questions. Each question consists of six randomly ordered results, generated by VDSR, DnCNN, PyramidCNN, U-Net, V\_Concate and our method, on a randomly selected test image. 60 participants have to choose 1 to 2 images that they perceive most appealing and comfortable. After averaging the votes from all the 20 questions, we obtain the statistics in Fig. 14. It is clear that the proposed model is more preferable to the human visual system, although U-Net and V\_Concate achieve high scores under certain numerical image quality measures.

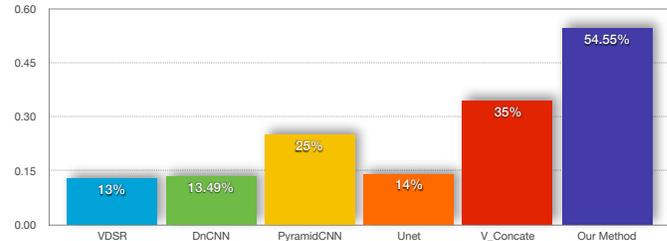


Fig. 14. User study on moiré pattern restoration.

## VII. MODEL VERSATILITY

### A. Cross-Data Evaluation

We quantitatively measure our model versatility by training and testing on data collected with different phone models or digital monitors. We perform three experiments, including testing on images taken with an iPhone on a Mac 2560 screen, with a SamSung S7 on a Dell 1920 monitor, and with a Sony Z5 on a Dell 1280 display, respectively. Note that in each experiment, the test data is excluded during training process. The performance is demonstrated in Table. VII. Though the performance is not as good as before, our model can still produce reasonable results. We also observe that the quality improvement by our model is most noticeable when the input (moiré) images are in low quality, such as the images captured with the Sony Z5 on the DELL 1280 screen.

TABLE VII  
CROSS-DATA EVALUATION.

Test Data\Metrics	PSNR		SSIM		FSIM	
	Input	Result	Input	Result	Input	Result
iPhone_Mac2560	23.09	25.18	0.840	0.862	0.914	0.930
SamSung_Dell1920	18.34	20.84	0.594	0.636	0.833	0.870
Sony_Dell1280	16.33	23.28	0.706	0.822	0.856	0.898

a) *Test on phone model HUAWEI P9:* Though the camera sensors in different phone models are different, the underlying reason for the formation of moiré patterns is similar on different phones. To test the versatility of our network, we run our network directly on moiré photos captured by another phone model, that is not used in collecting our dataset, HUAWEI P9. Decent results have been achieved, as shown in Fig. 15 and the supplemental materials. This indicates that our trained network can be used for removing moiré patterns in images captured by other phone models.



Fig. 15. Restoration of moiré photos taken with HUAWEI P9. Our model is not fine-tuned for this phone model.

**B. Restore Partial Moiré Photos**

*a) Synthesised moiré images:* Moiré patterns on an image can be spatially varying, strong in a region and weak in another region. Under extreme conditions, moiré patterns can only appear in part of an image. In Fig. 16, we show our results on synthesised partial moiré images, where only a small portion of the image contains moiré artefacts.



Fig. 16. Test on synthesised images contaminated by moiré patterns in a small region.

*b) Real world moiré patterns not caused by display:* When searching the Internet for “moiré photos”, we find that moiré patterns most commonly appear on fine repetitive patterns, such as textile textures on clothes and buildings. In Fig. 17, we show the results of directly applying our trained model without fine-tuning on Internet images damaged by moiré artefacts. Though the moiré is caused by the repetition of the fine patterns rather than digital display, our model is able to reduce such moiré patterns as well.

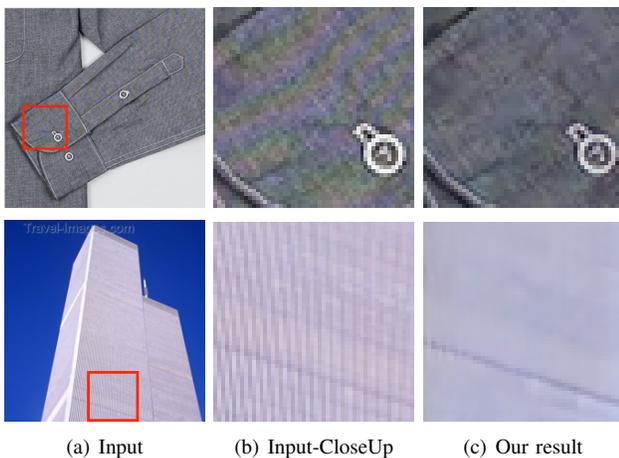


Fig. 17. Reduce moiré artefacts on Internet images without fine-tuning. Image courtesy @Fstoppers user Peter House and @Travel-Images.com user A.Bartel, respectively.

**VIII. LIMITATIONS**

When a moiré pattern exhibits very severe large-scale coloured bands, our method might not be able to infer the uncontaminated image correctly. We show a failure case in Fig. 18.

Another limitation is that our model could not clearly reduce blurriness in the input images. Note that other baseline algorithms, including the image deblurring model PyramidCNN, are not able to resolve it either (Fig. 12). We believe that such blurriness is introduced into a subset of acquired photos in our dataset because of multiple reasons, including motion blur due to the movement of the camera during image acquisition, the imperfect image alignment during pre-processing, and the damaged high-frequency components caused by high-frequency moiré patterns. Although our algorithm can faithfully detect all 20 corner points, moiré patterns can interfere with their exact localisation, giving rise to imperfect alignment.

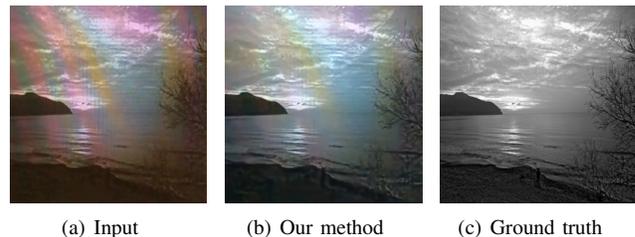


Fig. 18. A failure example.

**IX. CONCLUSION AND FUTURE WORK**

To conclude, we presented a novel multiresolution fully convolutional network for automatically removing moiré patterns from photos as well as created a large-scale benchmark with 100,000+ image pairs to evaluate moiré pattern removal algorithms. Although a moiré pattern can span over a wide range of frequencies, our proposed network is able to remove moiré artefacts within every frequency band thanks to the non-linear multiresolution analysis of the moiré photos. We believe that people would like to use their mobile phones to record content on screens for more reasons than expected, such as convenience, simplicity, and efficiency. The proposed method and the collected large-scale benchmark together provide a decent solution to the moiré photo restoration problem.

In the future, we would like to explore different categories of moiré patterns and improve our method so that it can eliminate moiré artefacts according to their category labels. Moreover, it will be interesting to investigate the existence of an indicator that can better describe the level of moiré artefacts and guide the training process. We also plan to keep expanding our dataset by adding more examples under different shooting conditions and for different types of device screens. We believe that with a larger dataset, our method can produce even better results.

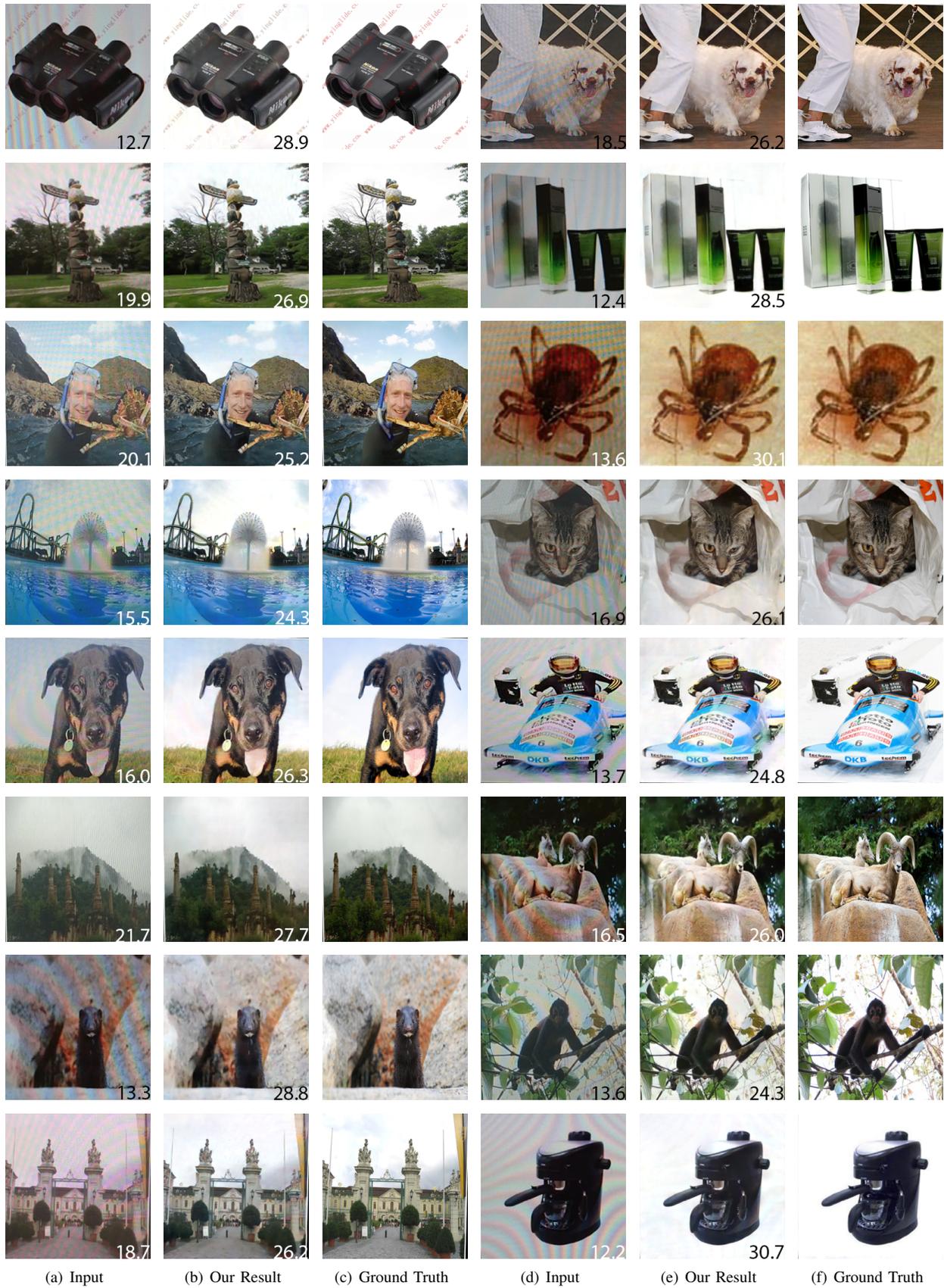


Fig. 19. Input images contaminated with different types of moiré patterns and their corresponding cleaned results from our proposed method. In this figure, we intentionally show some brighter images, where moiré patterns are more noticeable.

## ACKNOWLEDGEMENTS.

This work was partially supported by Hong Kong Research Grants Council under General Research Funds (HKU17209714).

## REFERENCES

- [1] X. Chen, S. Kang, J. Yang, and J. Yu, "Fast patch-based denoising using approximated patch geodesic paths," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 1211–1218.
- [2] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Transactions on Graphics (TOG)*, vol. 31, no. 6, p. 139, 2012.
- [3] H. Cho, H. Lee, H. Kang, and S. Lee, "Bilateral texture filtering," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 4, p. 128, 2014.
- [4] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [6] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European Conference on Computer Vision*. Springer, 2014, pp. 184–199.
- [7] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.
- [8] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 191, 2016.
- [9] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, 2017.
- [10] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," *CVPR*, 2017.
- [11] Wikipedia, "Moire pattern," 2017. [Online]. Available: [https://en.wikipedia.org/wiki/Moir%C3%A9\\_pattern](https://en.wikipedia.org/wiki/Moir%C3%A9_pattern)
- [12] KeohiHDTV, "Moire," 2017. [Online]. Available: <http://www.keohi.com/keohihdtv/learnabout/definitions/moire.html>
- [13] D. N. Sidorov and A. C. Kokaram, "Suppression of moiré patterns via spectral analysis," in *Proc. SPIE*, vol. 4671, 2002, p. 895.
- [14] F. Liu, J. Yang, and H. Yue, "Moiré pattern removal from texture images via low-rank and sparse matrix decomposition," in *Visual Communications and Image Processing (VCIP)*, 2015. IEEE, 2015, pp. 1–4.
- [15] F. Sur and M. Grediac, "Automated removal of quasiperiodic noise using frequency domain statistics," *Journal of Electronic Imaging*, vol. 24, no. 1, pp. 013 003–013 003, 2015.
- [16] N. Damera-Venkata and B. L. Evans, "Adaptive threshold modulation for error diffusion halftoning," *IEEE Transactions on Image Processing*, vol. 10, no. 1, pp. 104–116, 2001.
- [17] Z. He and C. A. Bouman, "Am/fm halftoning: digital halftoning through simultaneous modulation of dot size and dot density," *Journal of Electronic Imaging*, vol. 13, no. 2, pp. 286–302, 2004.
- [18] P. W. Wong, "Inverse halftoning and kernel estimation for error diffusion," *IEEE Transactions on Image Processing*, vol. 4, no. 4, pp. 486–498, 1995.
- [19] H. Siddiqui and C. A. Bouman, "Training-based descreening," *IEEE transactions on image processing*, vol. 16, no. 3, pp. 789–802, 2007.
- [20] Y.-W. Shou and C.-T. Lin, "Image descreening by ga-cnn-based texture classification," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 51, no. 11, pp. 2287–2299, 2004.
- [21] J. Luo, R. De Queiroz, and Z. Fan, "A robust technique for image descreening based on the wavelet transform," *IEEE Transactions on Signal Processing*, vol. 46, no. 4, pp. 1179–1184, 1998.
- [22] H. Siddiqui, M. Boutin, and C. A. Bouman, "Hardware-friendly descreening," *IEEE Transactions on Image Processing*, vol. 19, no. 3, pp. 746–757, 2010.
- [23] B. Sun, S. Li, and J. Sun, "Scanned image descreening with image redundancy and adaptive filtering," *IEEE Transactions on Image Processing*, vol. 23, no. 8, pp. 3698–3710, 2014.
- [24] J. Ok, S. Youn, G. Seo, E. Choi, Y. Baek, and C. Lee, "Paper check image quality enhancement with moire reduction," *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21 423–21 450, 2017.
- [25] L. Karacan, E. Erdem, and A. Erdem, "Structure-preserving image smoothing via region covariances," *ACM Transactions on Graphics (TOG)*, vol. 32, no. 6, p. 176, 2013.
- [26] S. Ono, T. Miyata, and I. Yamada, "Cartoon-texture image decomposition using blockwise low-rank texture characterization," *IEEE Transactions on Image Processing*, vol. 23, no. 3, pp. 1128–1142, 2014.
- [27] Y. Sun, S. Schaefer, and W. Wang, "Image structure retrieval via l0 minimization," *IEEE transactions on visualization and computer graphics*, 2017.
- [28] B. Ham, M. Cho, and J. Ponce, "Robust image filtering using joint static and dynamic guidance," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015.
- [29] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," *CVPR*, 2017.
- [30] C. Harris and M. Stephens, "A combined corner and edge detector." in *Alvey vision conference*, vol. 15, no. 50. Manchester, UK, 1988, pp. 10–5244.
- [31] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.
- [32] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [33] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "Fsim: A feature similarity index for image quality assessment," *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [34] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.