We thank the reviewers for the constructive and extremely helpful comments. We have updated the manuscript and believe it has substantially been improved. We appreciate that all reviewers have recognized our method's novelties by saying it is 'quite clever' (R1), 'a clever approach' (R2), 'very interesting, and potentially useful' (R4), and our effort to build new datasets (R1).

R1: Signal processing analysis. Thanks for your comment. We have described it in more detail in our new version. As explained in [11], moire patterns and noise are different; the former are prevalent more in low and mid-frequencies. DIP relies on the spectral bias of the CNN to learn lower frequencies first. So before DIP learns the high-frequency details of the image, moire patterns have appeared in the results of DIP. Mathematically, adding the moire image provides a good local minimum for the network to converge to. In fact, although the moire image is corrupted by moire patterns, it still retains many high-frequency details. The UNET can retain these high-frequency details in the iteration for the reconstruction of a clean image from the blur image that lacks high-frequency details.

R1: Real world examples. We did encounter the problems you mentioned in practice. To test on the real world examples, we did some preprocessing, e.g., alignment. The problems have been addressed in some papers, e.g., multi-focus image fusion methods [Pixel Convolutional Neural Network for Multi-Focus Image Fusion. Information Sciences, 2017; Multi-focus Image Fusion with a Deep Convolutional Neural Network. IJLEO, 2018] for image alignment, and depth estimation from focus methods [Depth estimation from focus and disparity. ICIP, 2016] to extract the moire area with the same depth. These methods can be applied in practice to put our model into cameras. However, our paper focuses on how to use focused and defocused image pairs to remove moire patterns.

R1: Metrics for the input images. The averages of PSNR/SSIM for the moire input in *SynScreenMoire* and *SynTextureMoire* are 23.08/0.809 and 22.73/0.784, respectively, which are much worse than the demoireing results by our method. Usually, screen demoireing is easier than texture demoireing because screen moire patterns appear throughout the image, whereas in texture demoireing, the moire patterns only appear in high-frequency areas which must be both identified and demoired.

R1: The 'K' ablation, and Section 3.3. A failure example has been shown in the 'K' of Fig. 3. Adding a fully-connected network to learn the blur kernel can make the image satisfy the total variation prior. Therefore, the full model can smooth the noise without damaging the original image content. We have updated Section 3.3 accordingly.

R1: Speed. Our method does not require multi-day training and large training data. But similar to other DIP-like methods, ours is slower than feed-forward networks. We will address this issue in our future work.

R2: Number of iterations. Our model does not have the problem of getting worse results when iterations are over some threshold, due to the constraint by the blur image. In all our experiments, the number of iterations is set to 3000 empirically.

R2: Other architectures. We also tried the encoder-decoder and ResNet. Their PSNRs are 0.11dB and 1.05dB, respectively lower than UNET. Many other image restoration methods have shown that UNET has advantages over the two structures.

R2: Limitation. The main limitation of the work is that the test-time training is slower than pre-trained networks.

R4: Premise, PSF. Although setting the PSF slightly larger than a pixel can prevent moire patterns, doing so will generate blurry images. We also note in practice, the PSF is a spatially variant function, and additionally depends on wavelength, so it not possible to engineer an ideal PSF that produces a uniformly sharp image without moire patterns. In contrast, our algorithm can produce clear and moire-free images.

R4: Significance in modern cameras. In the professional photography community, the use of optical low pass filters (OLPFs) as suggested by the reviewer is in fact controversial. Although OLPFs do greatly reduce moire patterns, they come at the expense of image sharpness, and increasingly manufacturers remove OLPFs to improve acutance. Mainstream companies (Nikon, Canon, Pentax, Sony) now all offer multiple DSLR cameras without OLPFs, e.g. Nikon D5600, D7500, D500. Most mainstream smartphone manufacturers (Apple, Google, Huawei) use Sony image sensors (e.g. IMX586, IMX363, etc.) which also do not include OLPFs instead relying on software to reduce moire effect. We confirmed this by producing moire patterns by photographing high frequency patterns using flagship mobile phones (Google Pixel 4, Huawei P40, OPPO Reno3). Further, we note the presence of moire artifacts is often used to justify lowered smartphone camera ratings as part of DXOMark reviews, further confirming the commercial significance of this problem. Please note, the SynTextureMoire dataset does not rely on photographing a screen and focuses on moire in high frequency textures.

R4: Evaluation: single vs two images. Please note, we *do* compare to other methods that also take two images as input, namely GF, DJF, SVLRM, and MSJF; please revisit Table 3 and Figure 4. We also compare to recent single image demoireing methods for comparisons to the SOTA. As this is the first paper to use both moire/blur images to demoire, there are no demoire-specific methods to compare to that accept these two inputs. However, based on the reviewer concern, we conducted a 'fair' experiment by feeding both the blur image and the moire image to the previous single image demoireing models for training and testing, and found doing so helps a little but still underperforms our proposed method. This result has been added to the revision.

R4: Deployment of our method. In Section 5.3, we describe how to easily apply our method to smartphones. In the typical capture mode, the camera applies an auto-focus algorithm, which can be modified to save an additional defocused image during the focusing process. In practice, it is not necessary to take an image with moire patterns and then take a defocused image.

R4: Relevance to NeurIPS. We note the list of NeurIPS 2020 Subject Areas includes "Computational Photography", and further our work is related to unsupervised learning. Given the prevalence of mainstream DSLR and smartphone cameras without OLPFs, the moire problem is both ubiquitous and pressing. We believe our solution is of interest to the NeurIPS community.

R4: Line 39-40. The low-pass filters mentioned in this paragraph are variable low-pass filters (line 38), where the word "variable" is missing. We have added it in the revision. Thank you for pointing this out.