1 Appendix

In this appendix, we give more analysis about our dataset and present more experimental results.

1.1 Moiré Differences in Raw Domain

Moiré shapes in R, G1, G2, and B channels are different. In Fig. 1, we present the zoom-in version of the moiré patterns in different raw channels for better observation. For the exampled image, its R channel only contains vertical and horizontal stripes with the same regular scale. In contrast, its G2 channel contains fine-scale moiré patterns besides the regular scale ones. To make this clear, we further present the DCT spectra of the four channels. The moiré patterns usually appear as nearly periodic stripes, which implies that the moiré patterns will be represented as strong peak spots in the DCT spectrum. As shown in the second row of Fig. 1, the R and G1 channels only have peak spots at the left-up corner, which represent coarse-scale moiré patterns and this is consistent with the observation in the image domain. Meanwhile, the G2 and B channels have peak spots at left-up, right-up, and left-down corners. It means that the two channels have both coarse-scale and fine-scale moiré patterns, which is consistent with the observation in the image domain. Note that, all the images are normalized (the DCT spectrum is also calculated from normalized image) to avoid the effects of different intensities of different CFA channels.

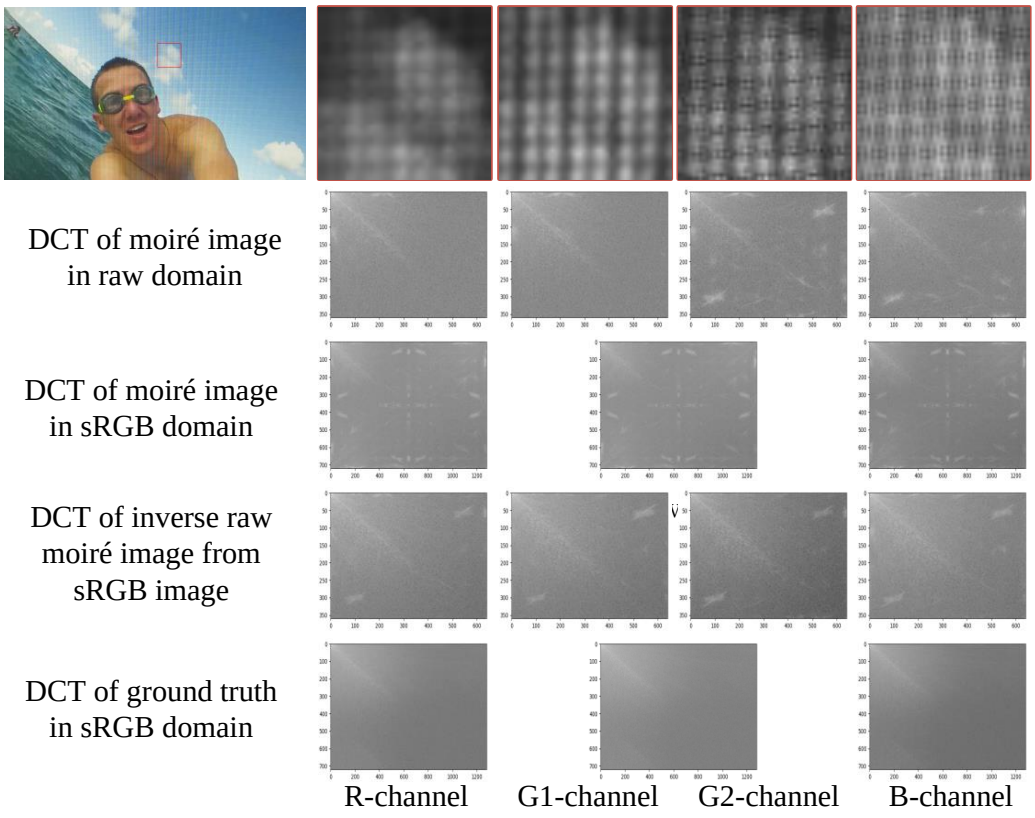


Figure 1: DCT spectrum of different channels.

We would like to point out that the moiré patterns are affected by the screen, camera, viewpoint, and image content. One possible reason for the moiré differences between these channels is that the viewpoints of the R, G1, G2, and B channels are different since there is a pixel shift in the CFA Bayer pattern. During data acquisition, even one-pixel shift can have a significant impact on the period and shape of moiré patterns. On the other hand, with the same camera, screen, and viewpoints but different contents, the generated moiré patterns are also different. For the R, G1, G2, and B channels of the same image, their contents are also different to some extent (due to different spectra responses of the color filter). Therefore, the moiré patterns of the four channels are different.

Meanwhile, the DCT spectrum of the R, G, and B channels of the RGB image are similar, as shown in the third row. The reason is that the demosaicing operation in the ISP process mixes these moiré patterns and the unique properties of one specific channel do not exist. Then we perform inverseISP (using the implementation provided by CycleISP) on the moiré RGB image, and the DCT spectrum of the inversed raw image are presented on the fourth row. It can be observed that they still have similar peak spots. In other words, the distinct moiré patterns in different raw channels cannot be recovered. Therefore, utilizing the real captured raw images for demoiréing network training is better than utilizing inversed (synthesized) raw images.

Our observation that different channels have different moiré patterns is representative. In Fig. 2, we provide eight examples of the moiré frames and give their DCT spectram. It can be observed that the DCT spectrum of different raw channels in the same image are different. This is consistent with our observations in the image domain that different raw channels have different moiré patterns.

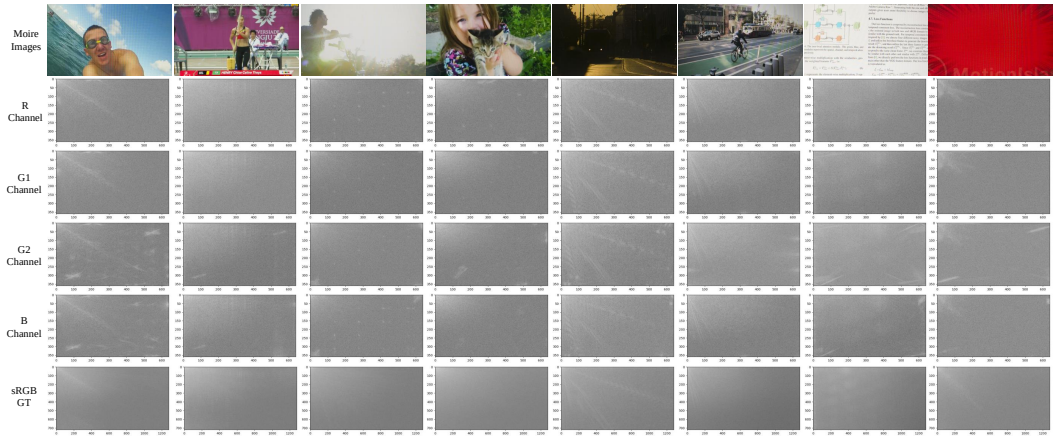


Figure 2: DCT spectra of different channels for eight examples.

1.2 Analysis About Our Raw Video Demoiréing Dataset

Diversity of source videos. Compared with the video demoiréing dataset in [1], the source videos in our dataset have a larger content diversity. Our dataset includes human activities, animal activities, landscapes, documents, GUIs, webpages, and animes. Fig. 3 presents some examples of our source videos and the percentages of different categories are also given.

Diversity of moiré patterns. The moiré patterns vary depending on the capturing distances, viewpoints, cameras, and display screens. In order to enrich the diversity of moiré patterns in our dataset, we utilize four different combines of cameras and screens, as shown in Table 1 for capturing. We also change the capturing distances and viewpoints to generate more diverse moiré patterns. As shown in Fig. 4 (a), the moiré patterns in our dataset have different scales, colors and stripes.

Temporal Confusion. As demonstrated in the main paper, it is difficult to keep the screen displaying and camera recording synchronization, which leads to temporal confusion problem (namely ghosting problem). Therefore, we developed an efficient temporal alignment method by inserting alternating patterns. In this way, the recaptured frames in our dataset are always clear, as shown in Fig. 4 (b). In contrast, the frames in [1] sometimes contain ghosting artifacts due to the temporal confusing problem, which will heavily affect the learning process.

Compare with synthesized moiré datasets. There is a significant gap between the synthesized moiré and real moiré patterns. Fig. 5 provides a comparison between CFAMoiré [7], LCDMoiré[6] and our real dataset. Various moiré pattern cycles can be observed on the same image for real captured ones. However, the synthesized moiré patterns lack this feature. In addition, real captured moiré patterns often have diverse colors and morphologies, which are relatively limited in synthesized moiré patterns. Finally, current moiré synthesis methods are all designed for sRGB domain moiré synthesis other than raw domain moiré synthesis.

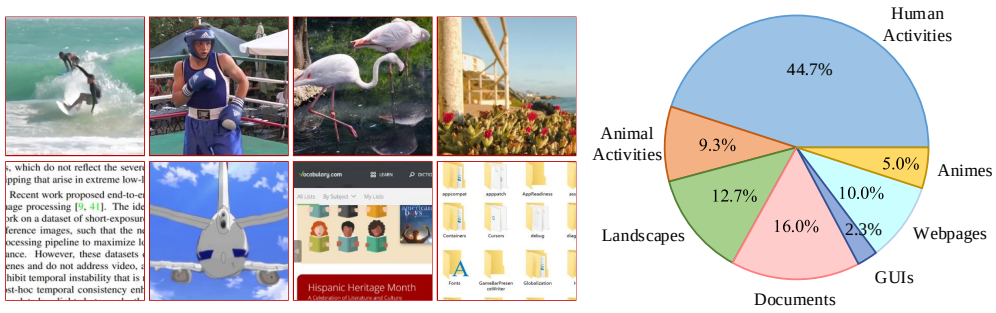


Figure 3: Examples of the source videos in our dataset (left) and the pie graph of categories (right).

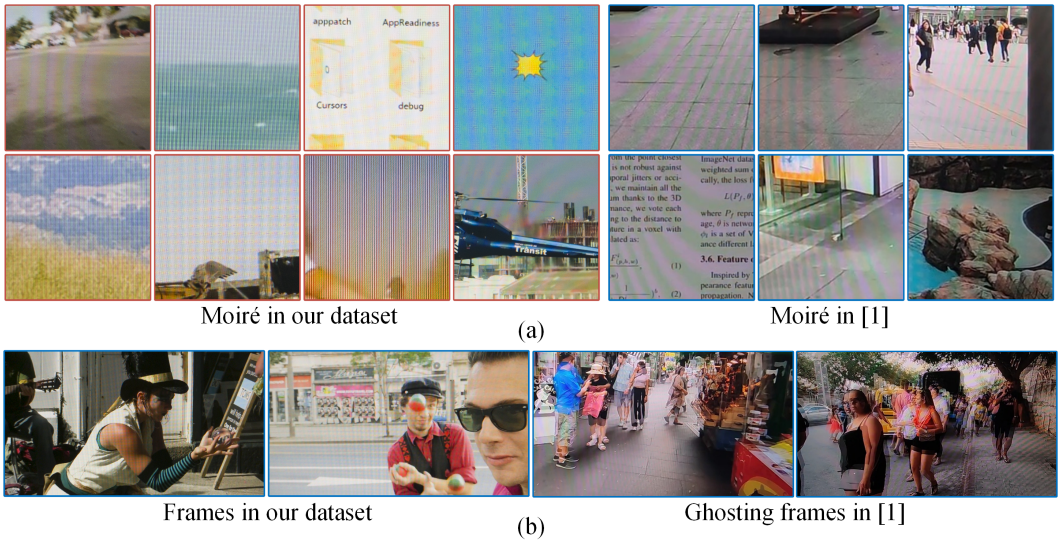
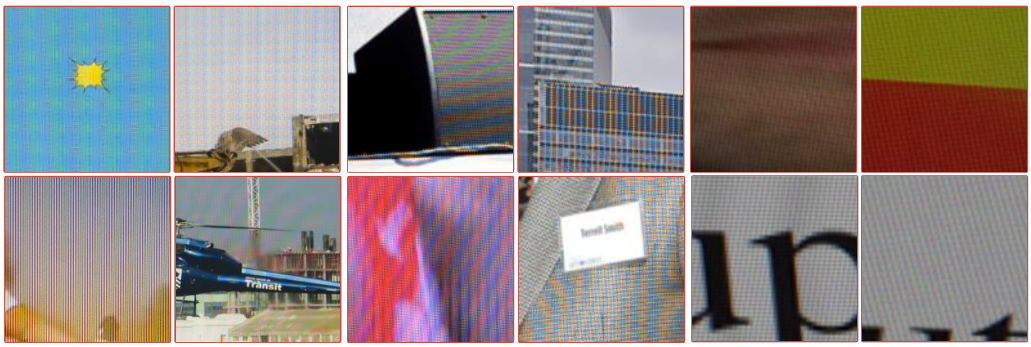


Figure 4: Comparison between our dataset and the dataset in [1]. (a) presents exampled moiré patterns in the two datasets and the moiré patterns in our dataset are more diverse. (b) illustrates the temporal confusion problem (i.e. generating ghosting frames) during recapturing, which is well solved by our method but existed in [1].

Table 1: The combinations of smart-phone cameras and screens used in our capturing process.

Phone	Screen	Screen Size (inch)	trainset	testset
Xiaomi	Dell G3 3590	15.6”	4020	720
Redmi	Lenovo Legion R7000P2021H	15.6”	4020	720
OnePlus 7	Lenovo Legion R9000P2021H	16”	3900	720
Honor V9	SKYWORTH 24X2	23.8”	3060	840



Our dataset CFAMoiré [27] LCDMoiré [28]

Figure 5: Examples of the moiré images in synthesized datasets and our dataset.

1.3 Ablation Study

1.3.1 Ablation on Channel Modulation

As shown in Figure 5 in the main paper, we utilize four color groups (R, G1, G2, and B) for channel modulation. In this experiment, we replace it with one color group but keep the channel number of extracted features be the same as that in color-separated features. As shown in Table 2, the three variants are all inferior to our complete version. This demonstrates that individual channels in the raw domain have limited features and are insufficient for effective fusion. In addition, our approach with GConv significantly reduces the parameter numbers.

Table 2: Ablation study results by exploring different channel modulation methods. The best results are highlighted in bold.

Variant	PSNR↑	SSIM↑	LPIPS↓
R-channel	28.559	0.9176	0.0932
G-channel	28.516	0.9182	0.0935
B-channel	28.551	0.9175	0.0914
Complete	**28.706**	**0.9201**	**0.0904**

1.3.2 Comparison of Modulation and Attention

Our "channel modulation" operation is different from the traditional "channel attention" operation. In Table 3, we perform an ablation experiment by replacing our channel modulation with traditional channel attention [2]. Despite channel attention utilizing more parameters (fully connected layers) to learn channel weights, its performance is inferior to that of our method. Fig. 6 presents the learned weights of channel modulation and channel attention. It can be observed that the channel attention weights are similar, which means there are no large differences. The main reason is that it needs a sigmoid layer to generate these coefficients. Different from it, our modulation coefficients are learned by setting them to be learnable parameters. The learned modulation coefficients have a large variance, which implies that different channels contribute very differently in the following.

Table 3: Comparison between our channel modulation and traditional channel attention. The best results are highlighted in bold.

Variant	PSNR↑	SSIM↑	LPIPS↓
Channel Attention	28.492	0.9170	0.0987
Channel Modulation	**28.706**	**0.9201**	**0.0904**

We also transform two raw image processing methods (RDNet [8] and UNet) for video demoiréing by incorporating the PCD module and fusion module. The results are presented in Table **??**. Our

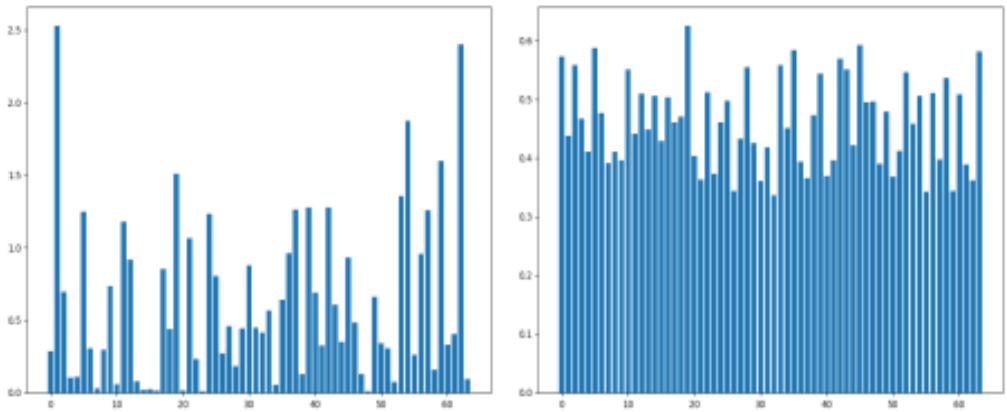


Figure 6: Comparison of learned weights between channel modulation and channel attention.

method consistently outperforms RDNet and UNet on all the three metrics. Despite RDNet being a raw domain approach, it did not consider the moiré differences in raw domain channels and it merely adjusted the channel number of the first convolution layer to fit the raw inputs, without any other tailored operations for raw data. In contrast, our method is tailored to the distinctive channel distribution properties of raw domain moiré patterns.

1.4 Temporal Consistency

We have provided a video demo to show the video moiré removal results and there are slightly jittering artifacts. To solve this problem, we further fine-tune our network by introducing the temporal loss proposed in [1]. After introducing temporal losses, our performance is further improved (see Table 4) and the temporal consistency is also improved.

Table 4: Comparison of with and without temporal loss. The best results are highlighted in bold.

Variant	PSNR↑	SSIM↑	LPIPS↓
w/o temporal loss	28.706	**0.9201**	0.0904
w temporal loss	**28.968**	0.9200	**0.0884**

1.5 Limitation

When the image has high color saturation and the moiré pattern contrast is strong, our method may cannot remove the patterns clearly, as shown in Fig. 7. Our method can remove the moiré patterns to some extent, while some compared methods failed in dealing with this hard example.

Figure 7: Demoiréing results in dealing with severe moiré patterns.

In terms of temporal consistency, we present additional results by fine-tuning with temporal loss presented in [1] in Table 4. We did not explore more advanced temporal consistency strategies, leaving this as a direction for future exploration.

2 Visual Comparisons for Image and Video Demoiréing

In this section, we present more visual comparison results for different demoiriéing methods. As introduced in the main paper, for image demoiréing, we present the visual comparison results with the first raw image demoiréing method RDNet [8] , and two state-of-the-art image demoiréing methods in sRGB domain, i.e., MBCNN [9] and UHDM [5]. We also present their results with raw inputs, denoted as MBCNN* and UHDM*. For video demoiréing, we present the visual comparison results for the six compared methods with raw inputs, including MBCNN*, UHDM*, VDMOiré* [1], EDVR* [4], and VRT^{-}* [3].

Figs. 8-9 present the image demoiréing results. It can be observed that RDNet and MBCNN cannot remove the moiré patterns clearly. UHDM also failed for the hard samples (as shown in Fig. 8). For the first image in Fig. 9, UHDM and UHDM* cannot recover the image details while removing moiré patterns. For the second image in Fig. 9, the compared methods tend to have color cast. In contrast, our method can recover the image details and remove the moiré patterns clearly.

Figs. 10-12 present the video demoiréing results. For the colorful scenes, the compared methods cannot remove the dense moiré patterns clearly, as shown in the first image of Fig. 10, Fig. 11, and the second image of Fig. 12. For image details, our method can recover the words in Fig. 12 clearly while the compared methods cannot recover them well. In summary, our method achieves the best performance in moiré removal and detail reconstruction. Note that, we also present a video demo to show the video demoiréing results.

References

[1] Peng Dai, Xin Yu, Lan Ma, Baoheng Zhang, Jia Li, Wenbo Li, Jiajun Shen, and Xiaojuan Qi. Video demoireing with relation-based temporal consistency. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17622–17631, 2022.

[2] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[3] Jingyun Liang, Jiezhang Cao, Yuchen Fan, Kai Zhang, Rakesh Ranjan, Yawei Li, Radu Timofte, and Luc Van Gool. Vrt: A video restoration transformer. *arXiv preprint arXiv:2201.12288*, 2022.

[4] Xintao Wang, Kelvin CK Chan, Ke Yu, Chao Dong, and Chen Change Loy. Edvr: Video restoration with enhanced deformable convolutional networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[5] Xin Yu, Peng Dai, Wenbo Li, Lan Ma, Jiajun Shen, Jia Li, and Xiaojuan Qi. Towards efficient and scale-robust ultra-high-definition image demoiréing. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XVIII*, pages 646–662. Springer, 2022.

[6] Shanxin Yuan, Radu Timofte, Ales Leonardis, and Gregory Slabaugh. NTIRE 2020 challenge on image demoiréing: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 460–461, 2020.

[7] Shanxin Yuan, Radu Timofte, Gregory Slabaugh, Aleš Leonardis, Bolun Zheng, Xin Ye, Xiang Tian, Yaowu Chen, Xi Cheng, Zhenyong Fu, et al. AIM 2019 challenge on image demoiréing: Methods and results. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3534–3545, 2019.

[8] Huanjing Yue, Yijia Cheng, Yan Mao, Cong Cao, and Jingyu Yang. Recaptured screen image demoiréing in raw domain. *IEEE Transactions on Multimedia*, 2022.

[9] Bolun Zheng, Shanxin Yuan, Gregory Slabaugh, and Ales Leonardis. Image demoiréing with learnable bandpass filters. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3636–3645, 2020.

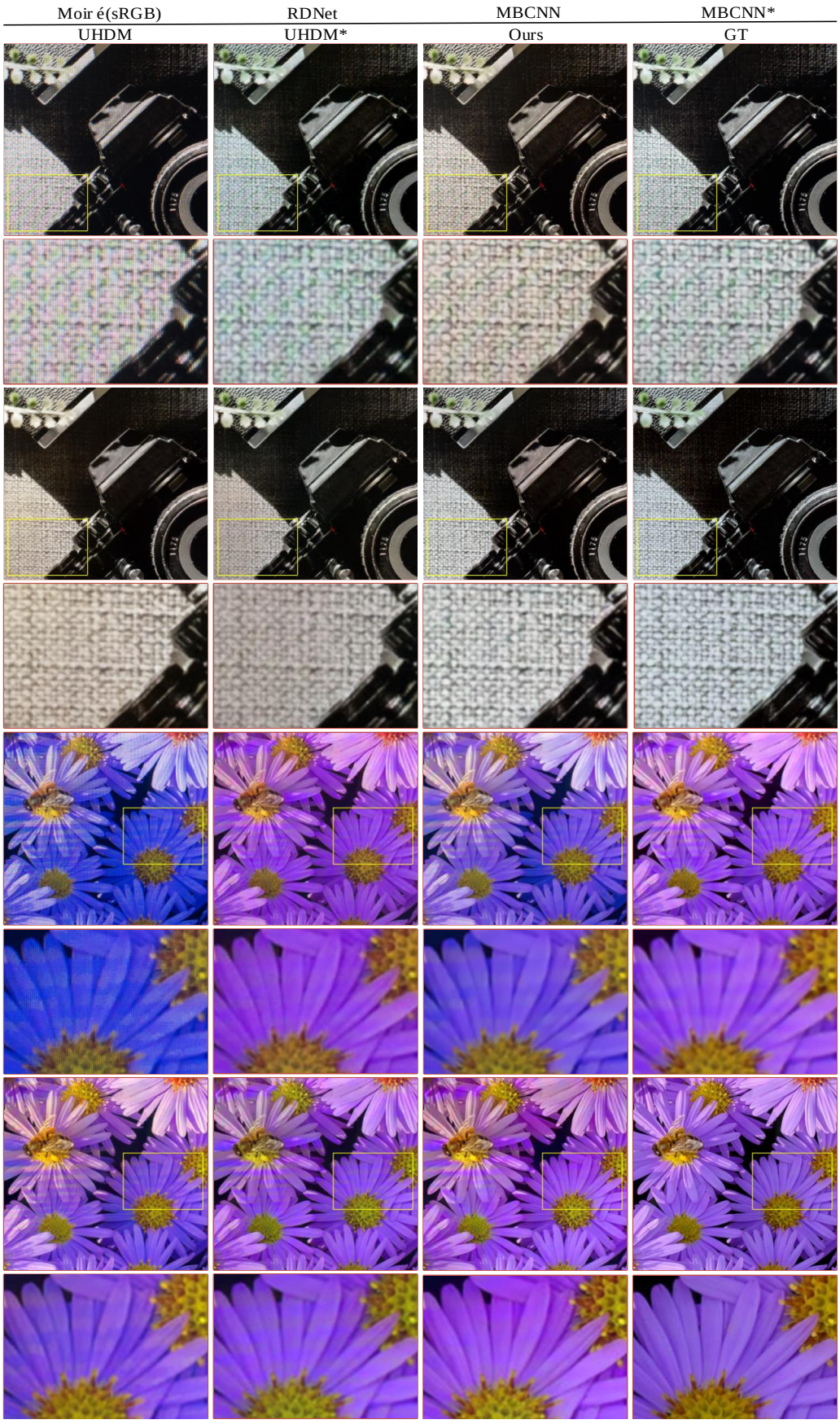


Figure 8: Visual comparison for image demoiréing. Zoom in for better observation.

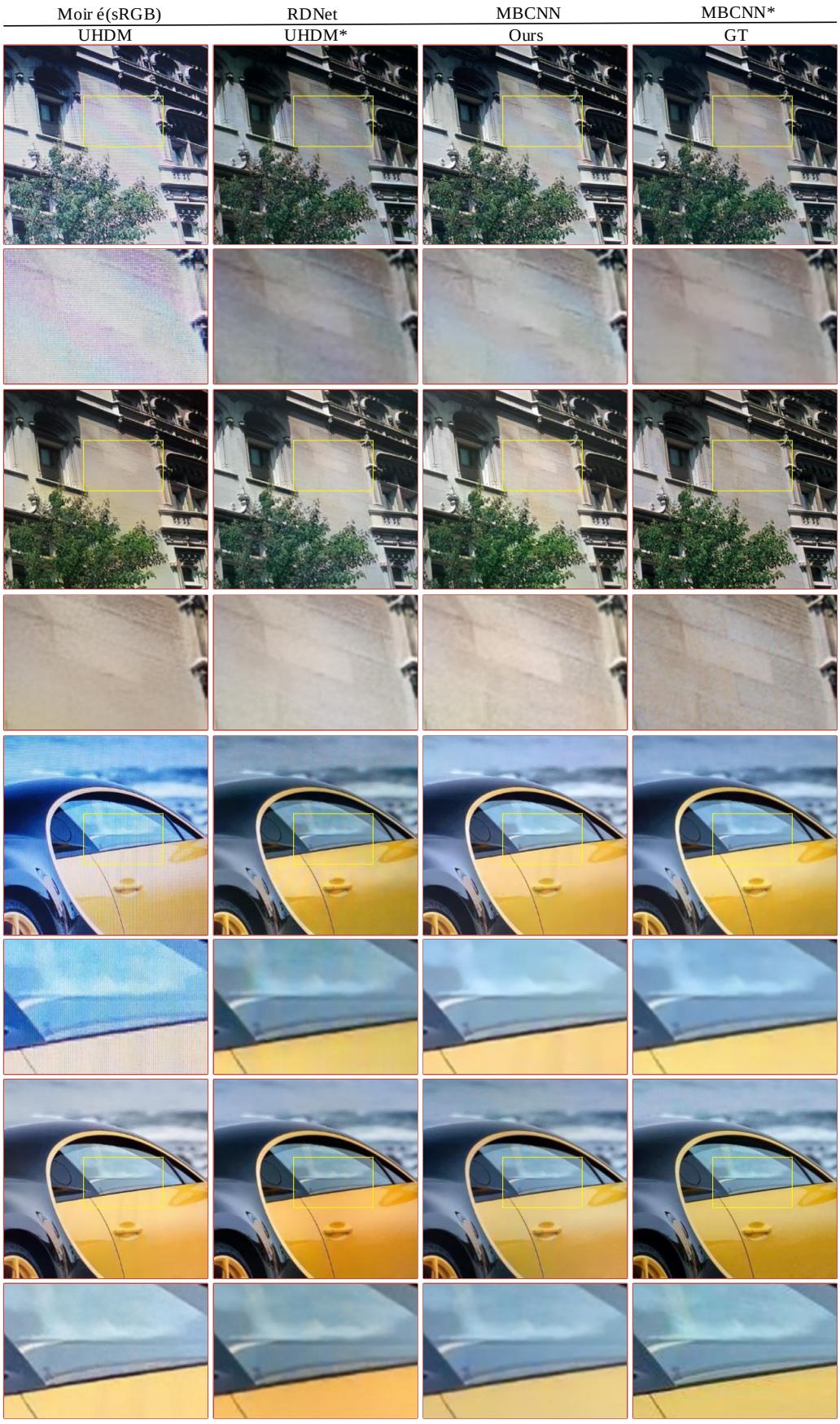


Figure 9: Visual comparison for image demoiréing. Zoom in for better observation.

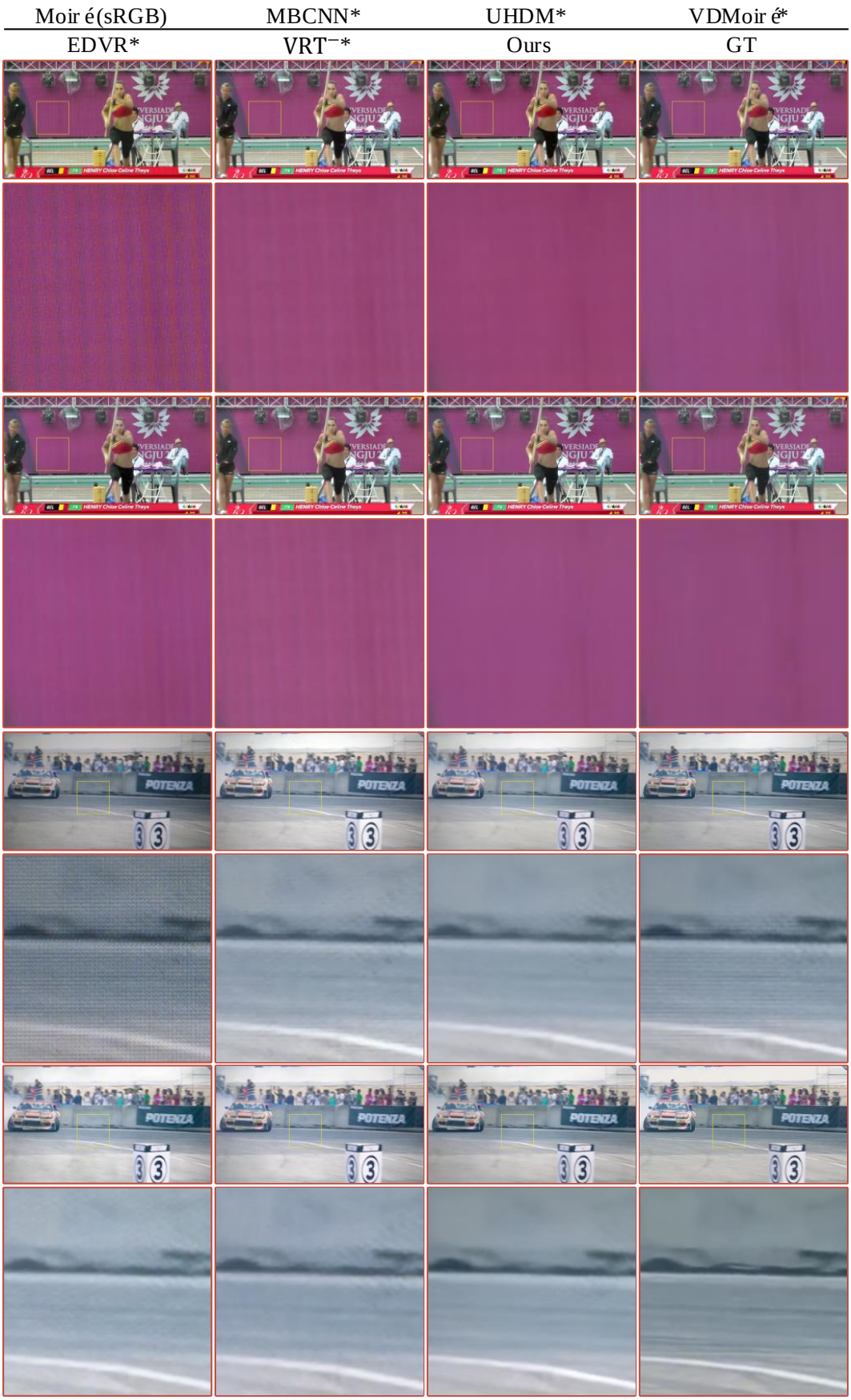


Figure 10: Visual comparison for video demoiréing. Zoom in for better observation.

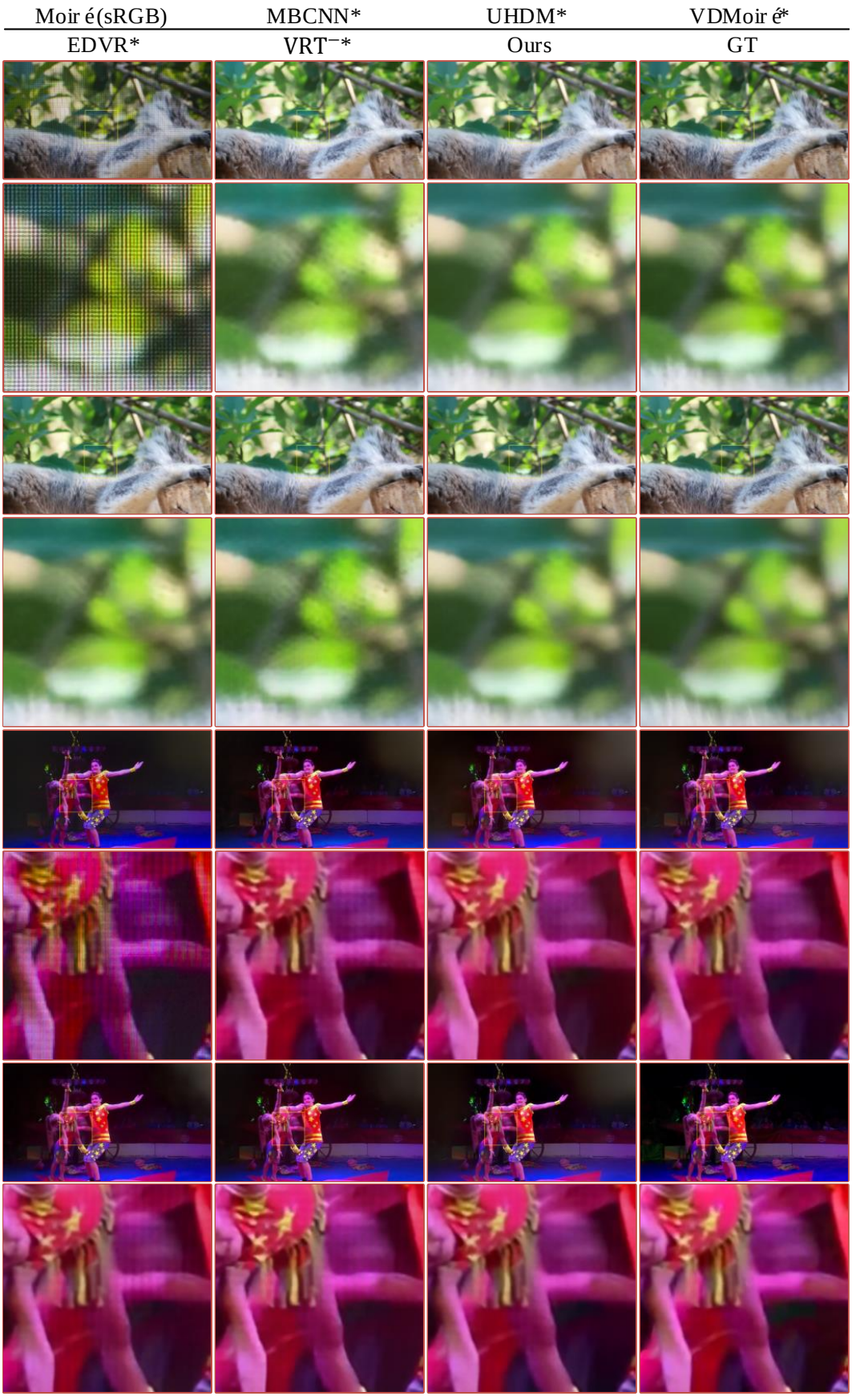


Figure 11: Visual comparison for video demoiréing. Zoom in for better observation.

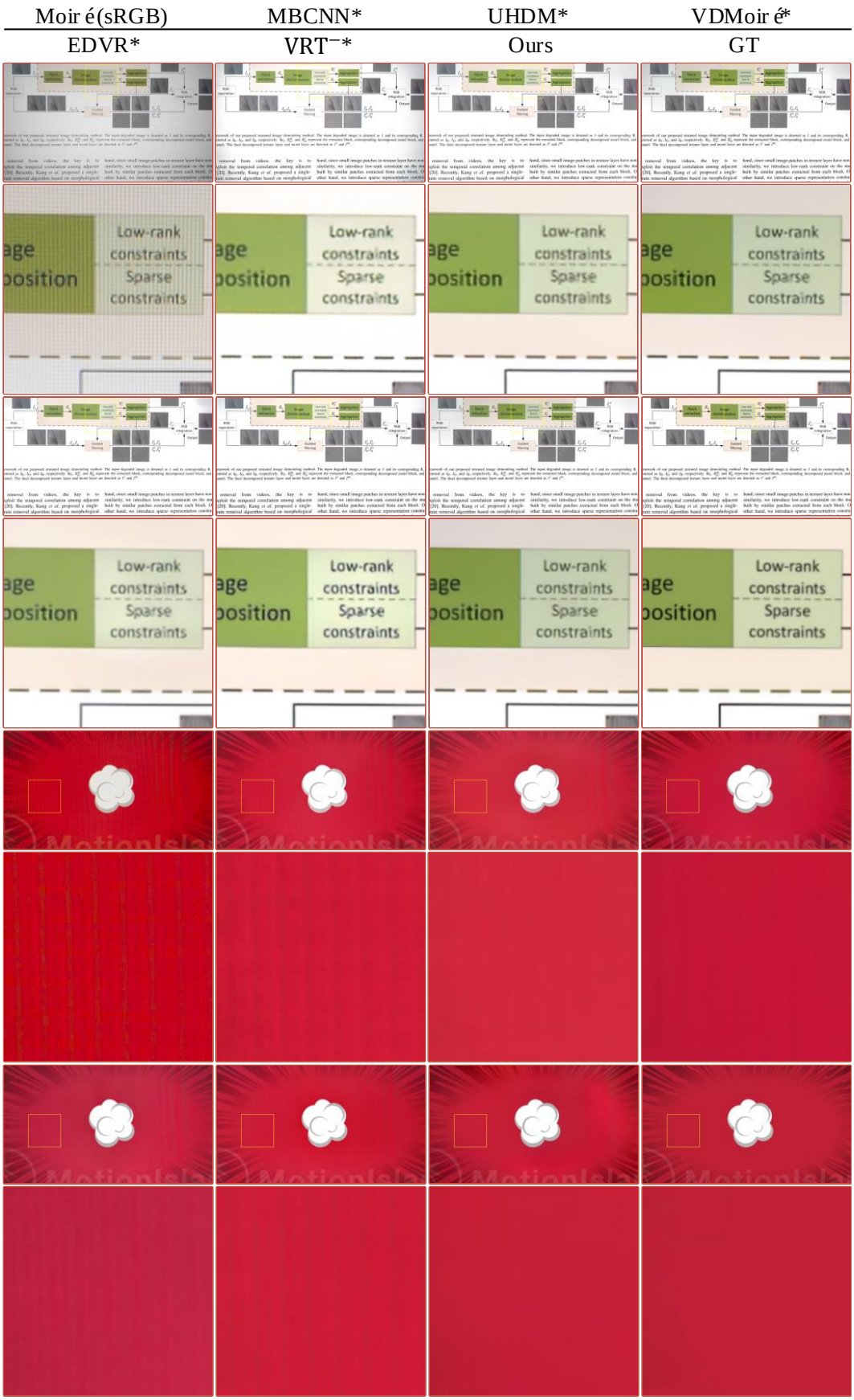


Figure 12: Visual comparison for video demoiréing. Zoom in for better observation.