Rolling Shutter Correction with Intermediate Distortion Flow Estimation
– *Supplementary Material* –

Mingdeng Cao[1] Sidi Yang[2] Yujiu Yang[2] Yinqiang Zheng[1]
[1]The University of Tokyo [2]Tsinghua University

1. More Ablation Studies

We further conduct ablation studies on our 3-frame-based model with the real-world rolling shutter correction dataset BS-RSC [1], following the same setting as the two-frame-based model depicted in the main paper.

Distortion flow estimation. The results of the models with different motion modeling settings are shown in Tab. 1, which is consistent with the 2-frame-based model shown in Table 3 of our main paper. Specifically, without motion modeling, the model achieves the lowest evaluation metrics, since it can hardly build the correlations between the latent GS frame and input RS frames. When with explicit modeling, the performance is improved significantly, demonstrating the significance and effectiveness of the explicit motion modeling for the high-quality RSC task. More specifically, the model with distortion flow estimation (and backward warping) surpasses the model with undistortion flow estimation (and forward warping) with a considerable performance improvement margin. Meanwhile, as the visual comparison shown in Fig. 1, we see that the model with distortion flow can better maintain the texture and obtain fewer artifacts. These results further demonstrate the effectiveness of the proposed distortion flow estimation with backward warping for the RSC task.

Flow attention and multi-flow decoder. The ablation results of the network designs (*i.e.*, flow attention mechanism and the multi-distortion flow decoding strategy) are shown in Tab. 2. The results are also consistent with the ablation results of the 2-frame-based model shown in our main submission. Specifically, the model without the flow attention mechanism exhibits a decline of 0.4dB on the PSNR. When adding the multi-distortion flow decoding strategy with 4 flow fields, the performance can be further improved by 0.1dB PSNR with a very slight runtime increment. However, as the number of flow fields continues to increase, the improvement in the metrics is limited (0.09dB PSNR), and there is a substantial increase regarding the runtime (an increment of 14ms). Moreover, as shown in Fig. 1, with more flow fields (other than 1 flow), the models can obtain more correct textures in the recovered GS image. As a result, we predict 4 distortion flow fields as default, obtaining a better trade-off between performance and efficiency.

Model	PSNR	SSIM	# Params.
W/o motion modeling	29.77	0.904	2.61
W/ undistortion flow $\mathbf{U}_{r\to g}$	31.38	0.925	3.05
W/ distortion flow $\mathbf{U}_{r\leftarrow g}$	32.40	0.938	2.98
Full model	34.48	0.954	3.15

Table 1. Effectiveness of the distortion flow estimation on the 3-frame-based model with real-world dataset BS-RSC [1]. The default setting of our final model is grayed.

Model	PSNR	SSIM	# Params.	Runtime
W/o Flow attention	33.94	0.950	2.99	30
W/ 1 field	34.38	0.952	3.14	31
W/ 4 fields	34.48	0.954	3.15	34
W/ 8 fields	34.57	0.954	3.16	48

Table 2. Effectiveness of the model designs on the 3-frame-based model with real-world dataset BS-RSC [1]. The default setting of our final model is grayed.

Method	BS-RSC		Fastec-RS	
	PSNR↑(dB)	SSIM↑	PSNR↑(dB)	SSIM↑
DRSC	24.97	0.872	23.86	0.818
CVR	24.58	0.823	26.67	0.838
Ours (2F)	<u>33.39</u>	<u>0.947</u>	<u>28.88</u>	<u>0.870</u>
Ours	**34.48**	**0.954**	**29.81**	**0.882**

Table 3. Additional quantitative comparison to CVR [2] and DRSC [7] on the real-world RSC dataset BS-RSC [1] and synthetic RSC dataset Fastec-RS [5].

2. Additional Comparison to the State-of-the-art

We further compare the proposed method to **1)** CVR [2], a method to recover the global shutter video from two consec-

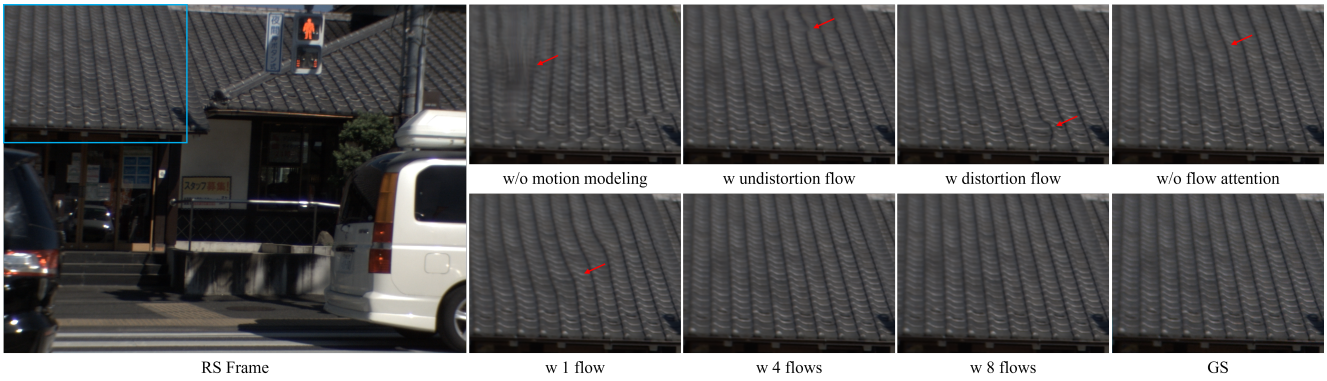

Figure 1. Visualization of ablation studies. We see that the full model (with 4 flows) with distortion flow estimation obtains a high-quality recovered GS image compared to the undistortion estimation-based method.

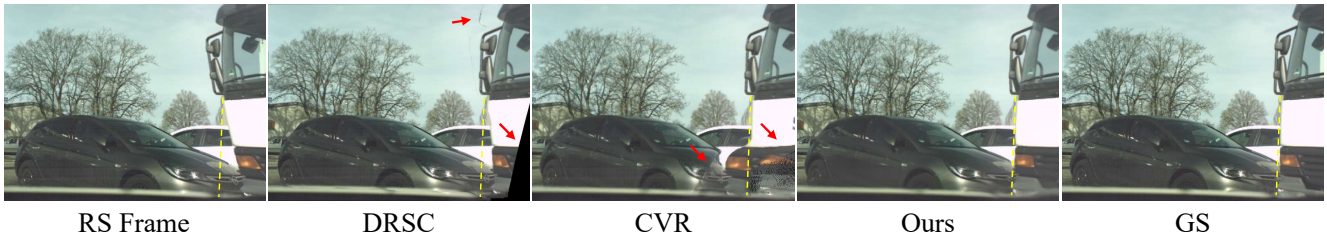

Figure 2. Additional qualitative comparison to CVR [2] and DRSC [7] on the synthetic RSC dataset Fastec-RS [5].

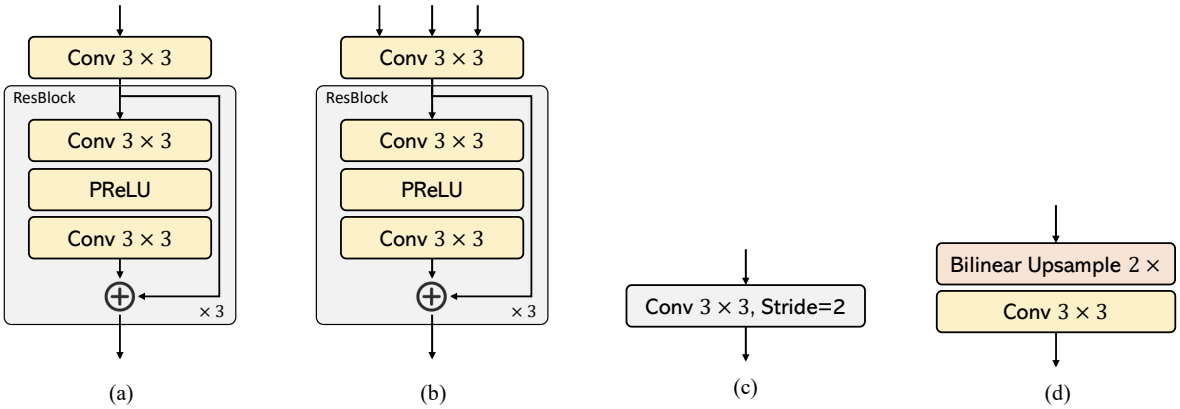

Figure 3. Network details. (a) Detailed structure of the `ConvBlock` in the image encoder. (b) Detailed structure of the `FusionConv` in the decoder part. (c) Convolution with stride 2 is used for downsampling. (d) Detailed structure of the upsampling module.

utive rolling shutter distorted frames, and **2)** DRSC [7] that directly scales the estimated optical flow for the correction. The quantitative and qualitative results are shown in Tab. 3 and Fig. 2. Our method outperforms them with a large performance margin regarding the PSNR and SSIM metrics. Meanwhile, as shown in Fig. 2, CVR and DRSC suffer from undesired artifacts (*e.g.*, black region, ghost effect), and our method performs well in this severely distorted scene.

3. Network Details

We here describe the detailed structure of the modules used in our main model architecture (shown in Figure 3 of our main submission).

`InConv`: A 7×7 convolutional layer that transfers the input RS images to feature maps.

`ConvBlock`: First, a 3×3 convolutional layer adjusts the channels of the input RS feature map, then three basic residual blocks [4] further extract the features. The detailed

structure is depicted in Fig. 3(a).

`DownSample`: A 3×3 convolutional layer with stride 2 is employed for downsampling.

`FusionConv`: The structure is similar to the `ConvBlock`. A 3×3 convolutional layer is employed to fuse the input GS feature map, warped RS feature maps, and the previous estimated distortion flows, then followed by three basic residual blocks [4]. The detailed structure is depicted in Fig. 3(b).

`UpSample`: We first apply the bilinear upsampling operation for 2 times upsampling and then a 3×3 convolutional layer is employed to output the upsampled distortion flow and GS feature.

`OutConv`: A 3×3 convolutional layer that transfers the GS feature map into the GS image.

4. More Qualitative Results

We further present more rolling shutter correction results on the synthetic Carla-RS [5] and the real-world BS-RSC [1] datasets in Figs. 4, 5 and 6. We see that existing methods DSUN [5], JAMNet [3], and QRSC [6] struggle to obtain satisfying results due to the non-linear and large motions, and noticeable artifacts exist in the restored GS images (please see the regions marked by rectangles and the arrows for better comparison). In contrast, our method can obtain higher-quality GS images that are closest to the ground-truth GS images, thanks to the direct distortion flow estimation and the delicate network designs.

References

[1] Mingdeng Cao, Zhihang Zhong, Jiahao Wang, Yinqiang Zheng, and Yujiu Yang. Learning adaptive warping for real-world rolling shutter correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17785–17793, 2022. 1, 3, 5, 6

[2] Bin Fan, Yuchao Dai, Zhiyuan Zhang, Qi Liu, and Mingyi He. Context-aware video reconstruction for rolling shutter cameras. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17572–17582, 2022. 1, 2

[3] Bin Fan, Yuxin Mao, Yuchao Dai, Zhexiong Wan, and Qi Liu. Joint appearance and motion learning for efficient rolling shutter correction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5671–5681, 2023. 3

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 2, 3

[5] Peidong Liu, Zhaopeng Cui, Viktor Larsson, and Marc Pollefeys. Deep shutter unrolling network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5941–5949, 2020. 1, 2, 3, 4

[6] Delin Qu, Yizhen Lao, Zhigang Wang, Dong Wang, Bin Zhao, and Xuelong Li. Towards nonlinear-motion-aware and occlusion-robust rolling shutter correction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10680–10688, 2023. 3

[7] Delin Qu, Bangyan Liao, Huiqing Zhang, Omar Ait-Aider, and Yizhen Lao. Fast rolling shutter correction in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):11778–11795, 2023. 1, 2

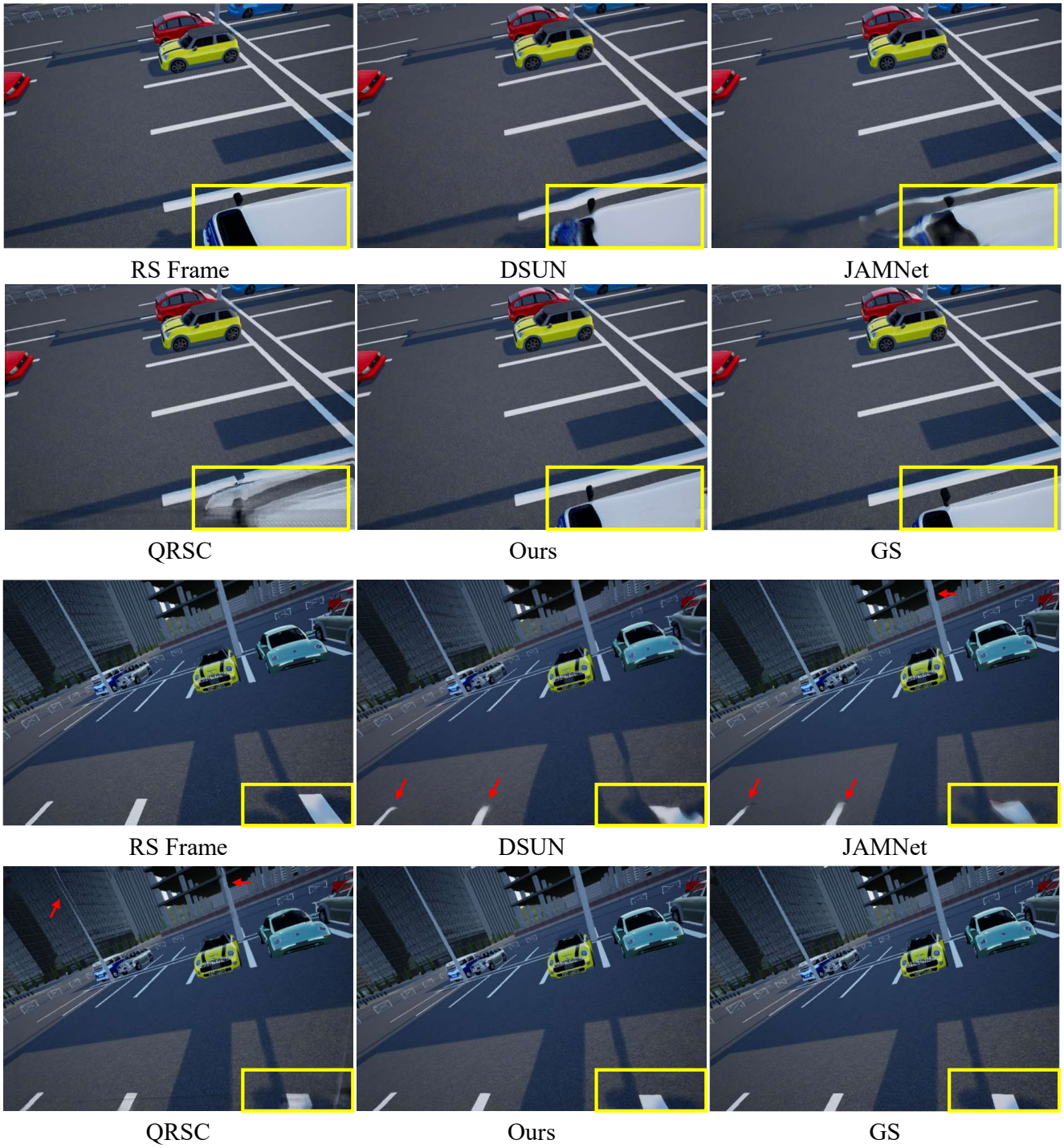

Figure 4. More qualitative results comparison against state-of-the-art methods on the synthetic Carla-RS dataset [5].

Figure 5. More qualitative results comparison against state-of-the-art methods on the real-world BS-RSC dataset [1].

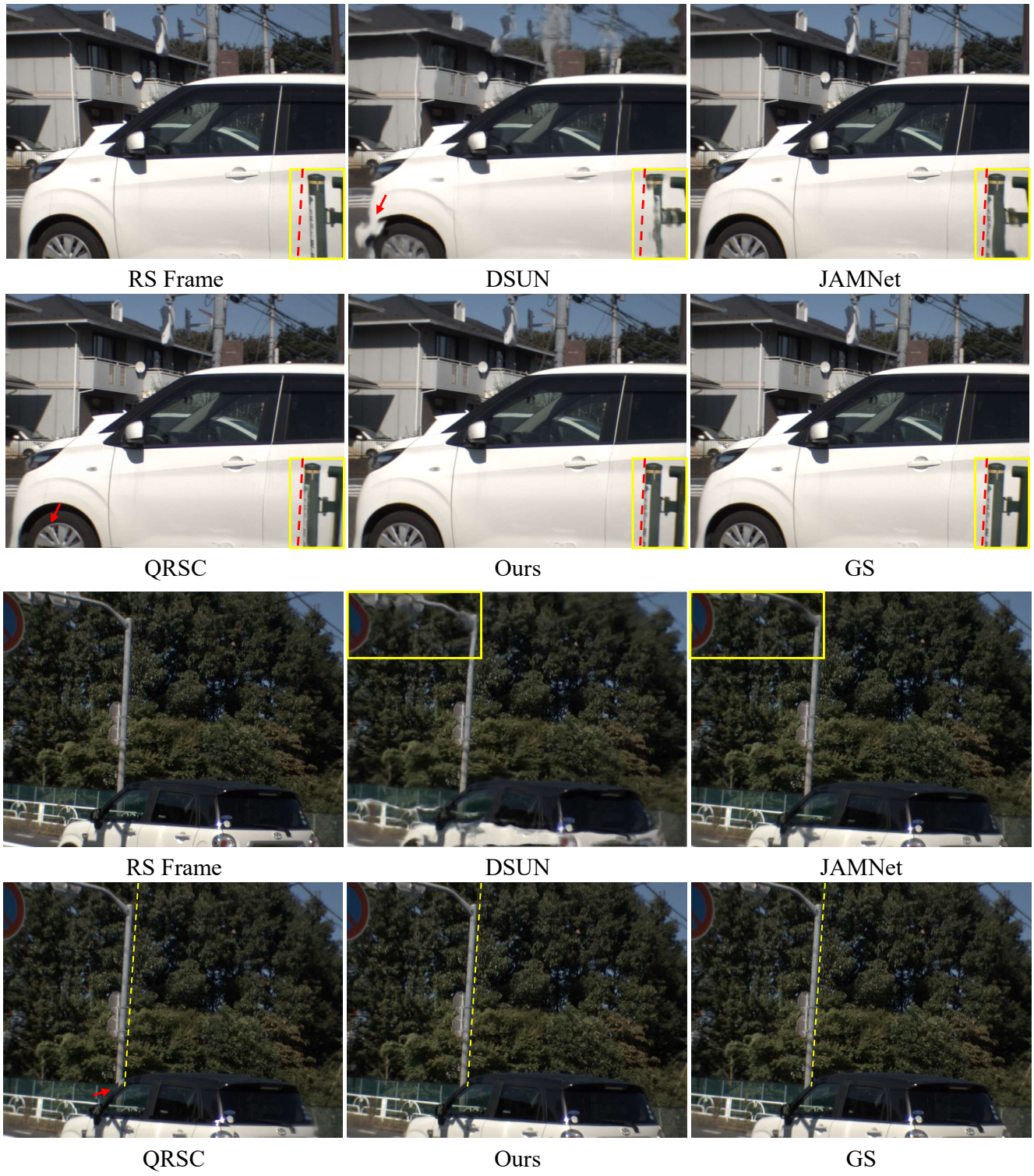

Figure 6. More qualitative results comparison against state-of-the-art methods on the real-world BS-RSC dataset [1].