Supplementary Material for: Efficient Event Stream Super-Resolution with Recursive Multi-Branch Fusion

Quanmin Liang[1,2*], **Zhilin Huang**[2,3*], **Xiawu Zheng**[2], **Feidiao Yang**[2], **Jun Peng**[2], **Kai Huang**[1†], **Yonghong Tian**[2,4†]
[1]School of Computer Science and Engineering, Sun Yat-Sen University [2]Peng Cheng Laboratory
[3]Shenzhen International Graduate School, Tsinghua University [4]Peking University
liangqm5@mail2.sysu.edu.cn, {zerinhwang03, yhtian}@pku.edu.cn, zhengxiawu@xmu.edu.cn,
{yangf, pengj01}@pcl.ac.cn, huangk36@mail.sysu.edu.cn

1 Datasets and Training Settings

Details of Synthetic Datasets: For the synthetic datasets NFS-syn and RGB-syn, we first downsample the NFS dataset [Kiani Galoogahi *et al.*, 2017] and the RGB-DAVIS dataset [Wang *et al.*, 2020] using bicubic interpolation to obtain corresponding images at different scales. The original resolution of the NFS dataset is 1280×720, and we perform $2(4, 8, 16)\times$ downsampling. The original resolution of the RGB-DAVIS dataset is 1520×1440, and we perform $2(4, 8)\times$ downsampling. Subsequently, the downsampled images are transformed into event streams using an event simulator [Lin *et al.*, 2022], following the default simulator parameters. Finally, we obtain the NFS-syn and RGB-syn datasets. NFS-syn has a minimum resolution of 80×45 with LR-HR pairs at $2(4, 8)\times$ scale factors. RGB-syn has a minimum resolution of 190×180 with LR-HR pairs at $2(4)\times$ scale factors. 80% of NFS-syn sequences are randomly selected for training all models, followed by validation on NFS-syn data and RGB-syn data.

Training Settings: To ensure fair comparison, we maintain the same training parameters as [Weng *et al.*, 2022]. Specifically, we set the learning rate to 10^{-4}, and decay it every 4000 iterations by a factor of 0.95. We utilize the ADAM optimizer and set the batch size to 2, training all methods for 100,000 iterations. All experiments are conducted on a Tesla V100 GPU.

2 Data Augmentation for ESR

For the Event Stream Super-Resolution (ESR) task, we employ the following methods for data augmentation:

- **Polarity flipping**: Flip all polarities p_i of event stream $\mathcal{E}$ with a probability of $p = 0.5$, i.e., $p_i = -p_i$.
- **RandomFlip** [Simonyan and Zisserman, 2015]: Flip low-resolution (LR) and high-resolution (HR) event images horizontally or vertically with a probability of $p = 0.5$.
- **Drop by time** [Gu *et al.*, 2021]: Discard all events in both LR and HR event streams within the same time period of $r_{time} \times \Delta_t$, where $r_{time} \in [0.1, 0.9]$, and Δ_t is the temporal length of the event stream.

[*]Equal Contribution
[†]Corresponding Author

Num	NFS-syn	RGB-syn	EventNFS
1	**0.300**	**0.0865**	**0.771**
2	0.303	0.0870	0.783
3	0.305	0.0877	0.788

Table 1: Comparison of choosing different numbers (Num) of data augmentation strategies from Selected DA. Training is conducted on the NFS-syn dataset, and $4\times$ SR testing is performed on NFS-syn, RGB-syn, and EventNFS datasets.

- **Random drop** [Gu *et al.*, 2021]: Randomly Drop a proportion $N \times p_{drop}$ of events in the LR event stream, where $p_{drop} \in [0.01, 0.09]$, and N represents the number of events in the LR event stream.
- **Drop by area** [Gu *et al.*, 2021]: Simultaneously discard a region $r_{cut}H \times r_{cut}W$ in both LR and HR event images, where $r_{cut} \in [0.05, 0.3]$, and H and W are the height and width of the LR or HR event images.
- **Random drop or add noise**: We propose to either drop a proportion $N \times p_{drop}$ of events or add noise events in the LR event stream with a probability $p_{drop} \in [-0.09, 0.09]$, where $p_{drop} < 0$ indicates event dropout, and $p_{drop} > 0$ indicates the addition of random noise.
- **Static Translation**: Shift both LR and HR event images vertically by $r_y \in [\pm 0.2 \times H]$ and horizontally by $r_x \in [\pm 0.2 \times W]$, where H and W are the height and width of the LR or HR event images.
- **RandomResizedCrop** [He *et al.*, 2016]: Randomly crop a region $r_{crop}H \times r_{crop}W$ from LR and HR event images and restore it to the original size using nearest interpolation, where $r_{crop} \in [0.25, 0.8]$.

We applied these data augmentation methods to the ESR task and compared their effectiveness. Additionally, inspired by RandAugment [Cubuk *et al.*, 2020], we combine Polarity flipping, RandomFlip, Drop by time, and Random drop or add noise into a data augmentation ensemble (Selected DA). To explore the optimal number of selections for Selected DA, we conducted comparative experiments. As shown in Table 1, selecting one data augmentation method randomly from Selected DA each time achieves the best results. This may

Method	NFS-syn	RGB-syn	EventNFS
EventZoom w/o DA	3.314	2.484	1.889
EventZoom Selected DA	1.021	1.117	1.354
RecEvSR w/o DA	0.374	0.364	0.811
RecEvSR Selected DA	**0.365**	**0.327**	**0.802**

Table 2: Comparison results of EventZoom and RecEvSR with and without using Selected DA.

be attributed to the sparse nature of event stream data, and applying too many data augmentations at once could excessively disrupt the spatial structure of event streams, leading to a performance decline.

Furthermore, to assess the effectiveness of our Selected DA, we conducted a comparison experiment on two other ESR methods, EventZoom [Duan *et al.*, 2021] and RecEvSR [Weng *et al.*, 2022], with and without the use of Selected DA. As shown in Table 2, the use of Selected DA significantly improves the performance and robustness of EventZoom and RecEvSR, reaffirming the efficacy of our approach.

Blocks Num	Param (M)	Time (ms)	NFS-syn	EventNFS
1	**2.0**	**5.8**	0.306	-
3	3.1	7.5	0.300	**0.316**
5	4.3	8.8	0.298	0.321
7	5.4	10.4	**0.294**	-

Table 3: Ablation study on the number of Residual Blocks in our RMFNET. The $4\times$ super-resolution results are presented on datasets NFS-syn and EventNFS.

Channel Size	Param (M)	Time (ms)	NFS-syn	EventNFS
64	**0.8**	**7.0**	0.314	0.334
128	3.1	7.5	0.300	0.316
256	12.1	12.1	**0.290**	**0.306**

Table 4: Ablation study on the size of feature channels in our RMFNET. The $4\times$ super-resolution results are presented on datasets NFS-syn and EventNFS.

3 Hyperparameters of Model

To validate the efficiency of our proposed Multi-Branch Information Fusion Network (RMFNET), we conducted a hyperparameter analysis on the number of Residual Blocks and the feature channel size (C) in the model. As shown in Table 3, increasing the number of Residual Blocks improves the performance of RMFNET on the NFS-syn dataset, with a gradual increase in parameters and inference time. However, it is observed that RMFNET encounters training instability issues on the EventNFS dataset when the number of blocks is set to 1 or 7. This might be attributed to the real dataset having a minimum resolution of 55×31 and severe degradation, requiring additional training techniques for model convergence when altering the depth.

With an increase in the size of feature channels, we notice a rapid growth in model parameters, accompanied by a significant improvement in performance (Table 4). It's worth noting that even with a model parameters of less than 1M ($C = 64$), our model's performance still surpasses that of the previous EventZoom [Duan *et al.*, 2021] and RecEvSR [Weng *et al.*, 2022], emphasizing the effectiveness of our proposed multi-branch architecture.

Therefore, in practical applications, a trade-off among inference speed, model parameters, and model performance can guide the selection of suitable model hyperparameters based on individual requirements.

4 Additional Visual Comparison Results

4.1 More Qualitative Comparison Results

As shown in Figure 1, Figure 2, and Figure 3, we present the $4\times$ super-resolution results of bicubic, SRFBN [Li *et al.*, 2019], RSTT [Geng *et al.*, 2022], EventZoom [Duan *et al.*, 2021], RecEvSR [Weng *et al.*, 2022], and our RMFNET on the NFS-syn, RGB-syn, and EventNFS datasets. For synthetic data, we trained on the NFS-syn dataset and tested on both NFS-syn and RGB-syn datasets. For real data, we randomly split it for training and testing. The results in the figures demonstrate that our RMFNET excels in recovering and complementing the missing details in the LR event stream, validating the effectiveness of our approach. In contrast, RecEvSR exhibits fragile generalization on RGB-syn, possibly due to the coordinate relocation it employs, making event images sparser and leading to overfitting during training, resulting in diminished model generalization. Conversely, our RMFNET demonstrates robust generalization, achieving the best visual results on RGB-syn, highlighting the stability of our approach.

4.2 Event-Based Video Reconstruction

As depicted in Figure 4, we conducted $4\times$ super-resolution on NFS-syn using bicubic, SRFBN [Li *et al.*, 2019], RSTT [Geng *et al.*, 2022], EventZoom [Duan *et al.*, 2021], RecEvSR [Weng *et al.*, 2022], and our RMFNET. Subsequently, we employed E2VID [Rebecq *et al.*, 2019] for video reconstruction based on these super-resolved event streams. It is evident that the reconstruction results corresponding to our RMFNET exhibit fewer artifacts and possess clearer edges and details, further confirming the effectiveness of our approach.

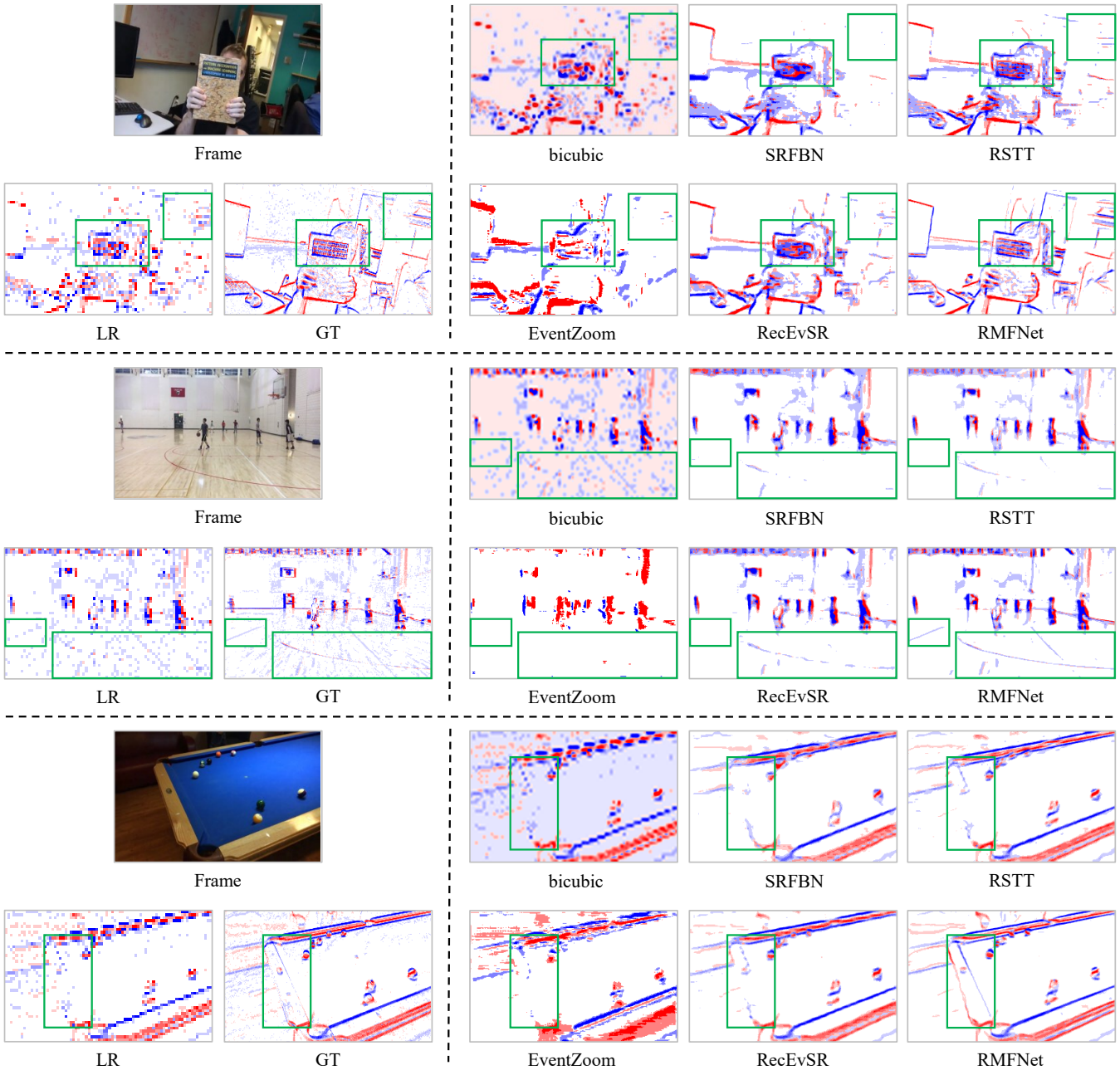

Figure 1: Qualitative comparison of $4\times$ super-resolution on NFS-syn dataset using bicubic, SRFBN, RSTT, EventZoom, RecEvSR, and our RMFNET. **Zoom in for the best view**.

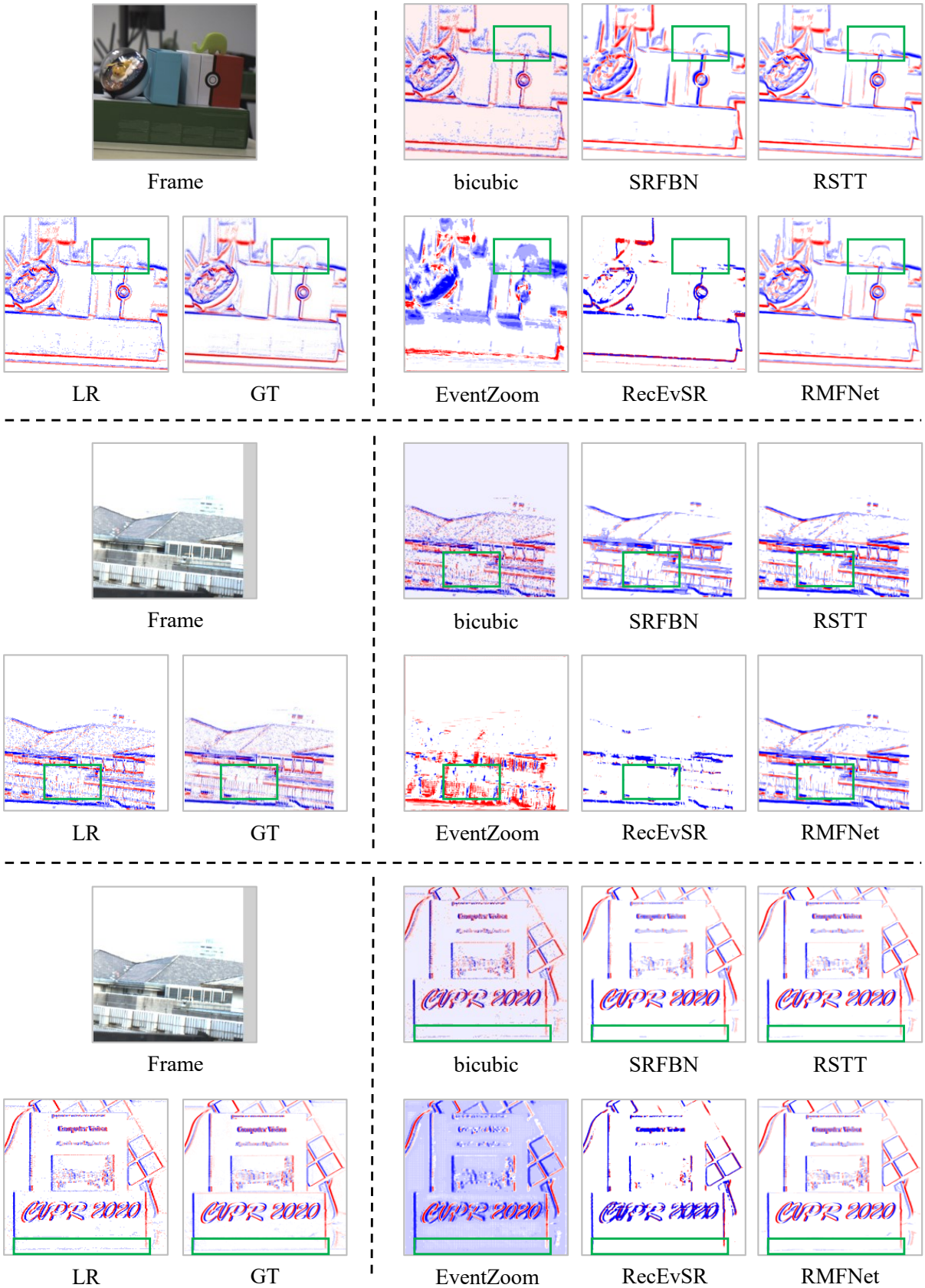

Figure 2: Qualitative comparison of $4\times$ super-resolution on RGB-syn dataset using bicubic, SRFBN, RSTT, EventZoom, RecEvSR, and our RMFNET. **Zoom in for the best view**.

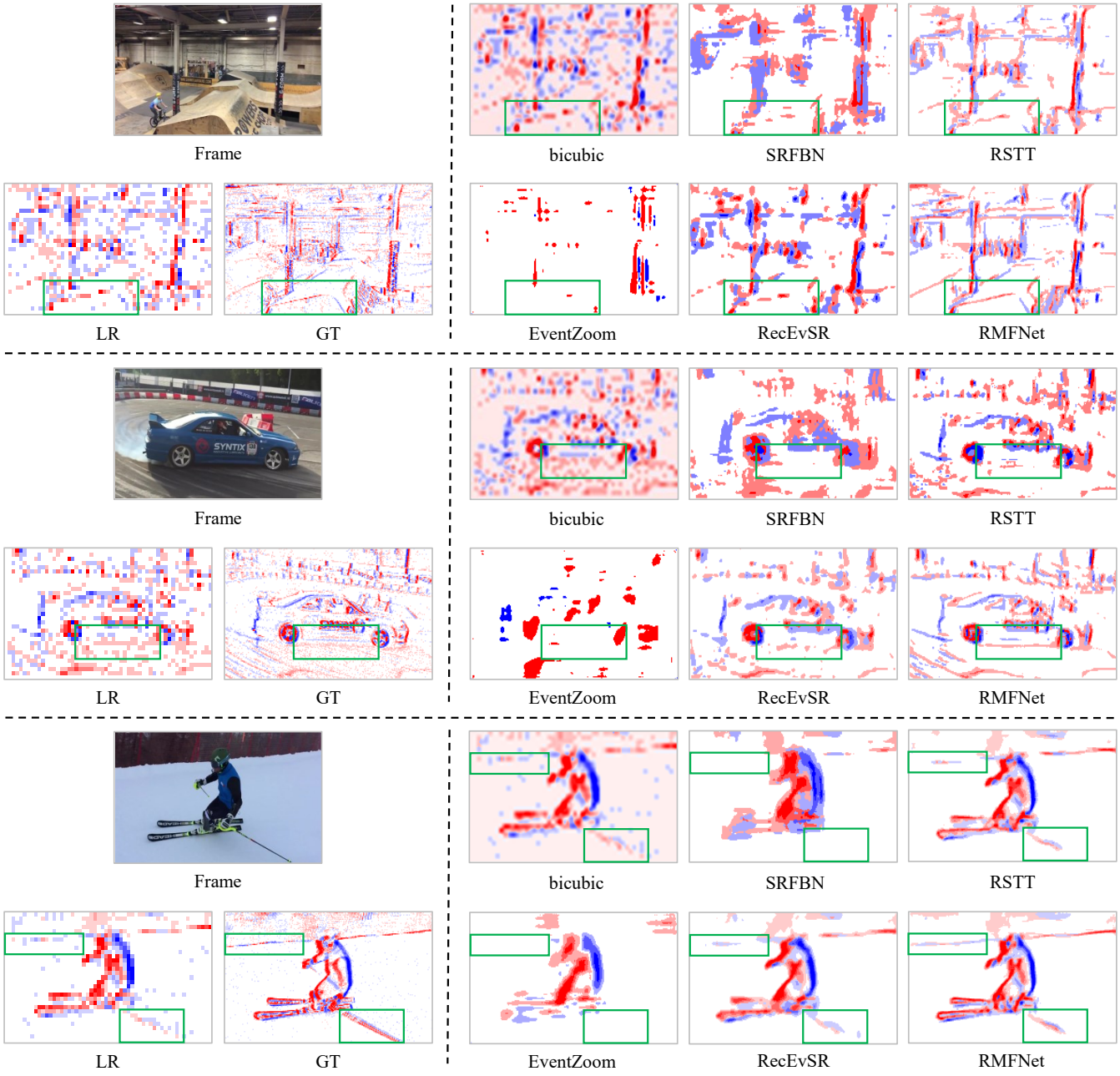

Figure 3: Qualitative comparison of $4\times$ super-resolution on EventNFS dataset using bicubic, SRFBN, RSTT, EventZoom, RecEvSR, and our RMFNET. **Zoom in for the best view**.

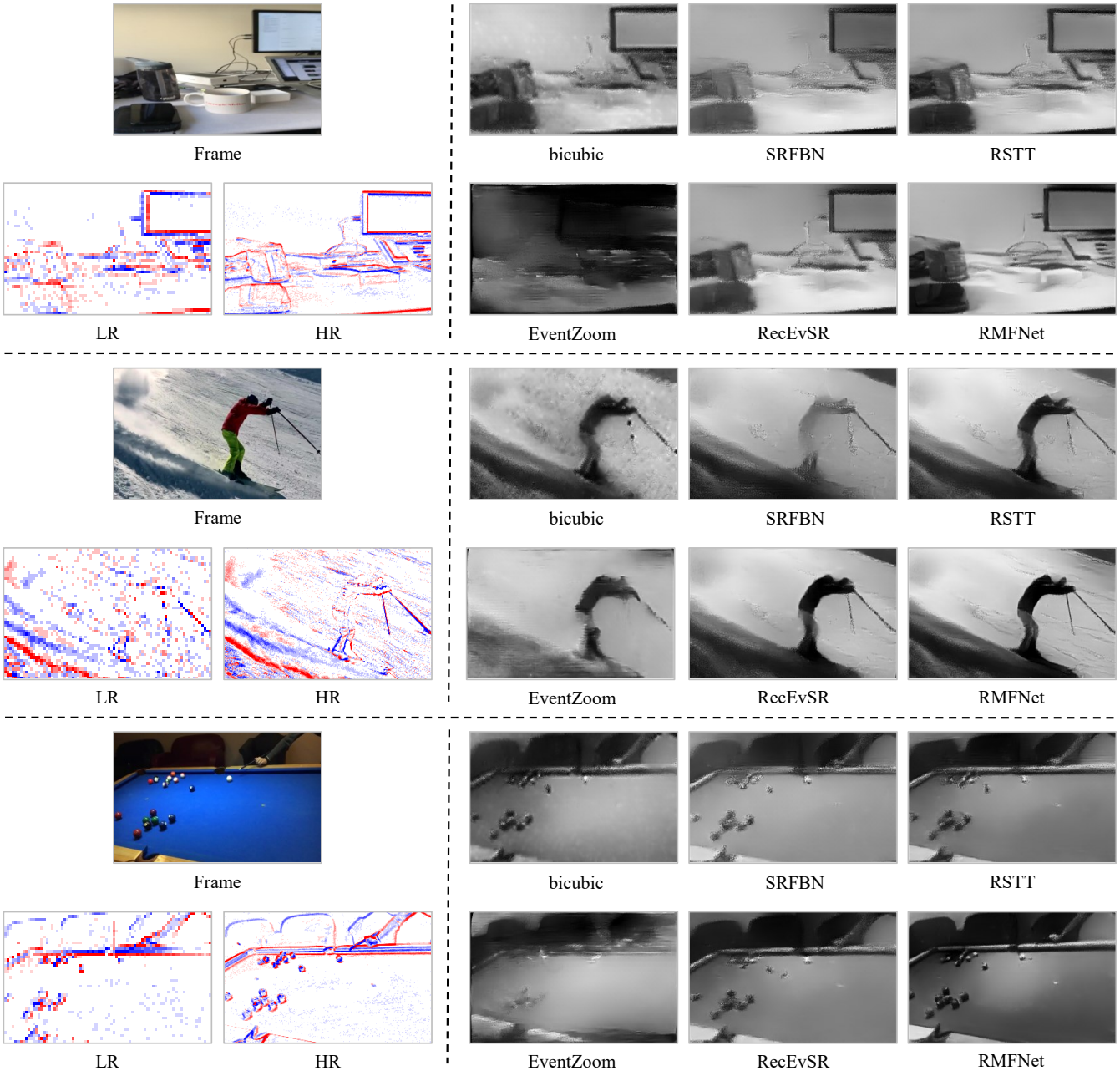

Figure 4: Qualitative comparison of image reconstruction based on $4\times$ event stream on the NFS-syn dataset.

References

[Cubuk *et al.*, 2020] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 702–703, 2020.

[Duan *et al.*, 2021] Peiqi Duan, Zihao W Wang, Xinyu Zhou, Yi Ma, and Boxin Shi. Eventzoom: Learning to denoise and super resolve neuromorphic events. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12824–12833, 2021.

[Geng *et al.*, 2022] Zhicheng Geng, Luming Liang, Tianyu Ding, and Ilya Zharkov. Rstt: Real-time spatial temporal transformer for space-time video super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17441–17451, 2022.

[Gu *et al.*, 2021] Fuqiang Gu, Weicong Sng, Xuke Hu, and Fangwen Yu. Eventdrop: Data augmentation for event-based learning. *arXiv preprint arXiv:2106.05836*, 2021.

[He *et al.*, 2016] 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.

[Kiani Galoogahi *et al.*, 2017] Hamed Kiani Galoogahi, Ashton Fagg, Chen Huang, Deva Ramanan, and Simon Lucey. Need for speed: A benchmark for higher frame rate object tracking. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1125–1134, 2017.

[Li *et al.*, 2019] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3867–3876, 2019.

[Lin *et al.*, 2022] Songnan Lin, Ye Ma, Zhenhua Guo, and Bihan Wen. Dvs-voltmeter: Stochastic process-based event simulator for dynamic vision sensors. In *European Conference on Computer Vision*, pages 578–593. Springer, 2022.

[Rebecq *et al.*, 2019] Henri Rebecq, René Ranftl, Vladlen Koltun, and Davide Scaramuzza. High speed and high dynamic range video with an event camera. *IEEE transactions on pattern analysis and machine intelligence*, 43(6):1964–1980, 2019.

[Simonyan and Zisserman, 2015] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In Yoshua Bengio and Yann LeCun, editors, *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015.

[Wang *et al.*, 2020] Zihao W Wang, Peiqi Duan, Oliver Cossairt, Aggelos Katsaggelos, Tiejun Huang, and Boxin Shi. Joint filtering of intensity images and neuromorphic events for high-resolution noise-robust imaging. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1609–1619, 2020.

[Weng *et al.*, 2022] Wenming Weng, Yueyi Zhang, and Zhiwei Xiong. Boosting event stream super-resolution with a recurrent neural network. In *European Conference on Computer Vision*, pages 470–488. Springer, 2022.