Supplementary Material for
GET: Group Event Transformer for Event-Based Vision

Yansong Peng Yueyi Zhang* Zhiwei Xiong Xiaoyan Sun Feng Wu
University of Science and Technology of China
pengyansong@mail.ustc.edu.cn, {zhyuey, zwxiong, sunxiaoyan, fengwu}@ustc.edu.cn

1. Code

The **code** is available at https://github.com/Peterande/GET-Group-Event-Transformer, containing the implementation of our GET network architecture.

Our code is structured in a modular and easy-to-follow manner, allowing for easy customization and extension. We include comments and explanations to aid understanding and reproducibility. The code is implemented using PyTorch 2.0.1 on Ubuntu 20.04 operation system.

2. Additional Visualizations

We provide additional visualizations of the object detection results on the Gen1 dataset, as shown in Figure 1. As can be seen, GET achieves better performance in detecting various objects in different scenarios, including objects that are distant, obscured, or have low illuminance changes.

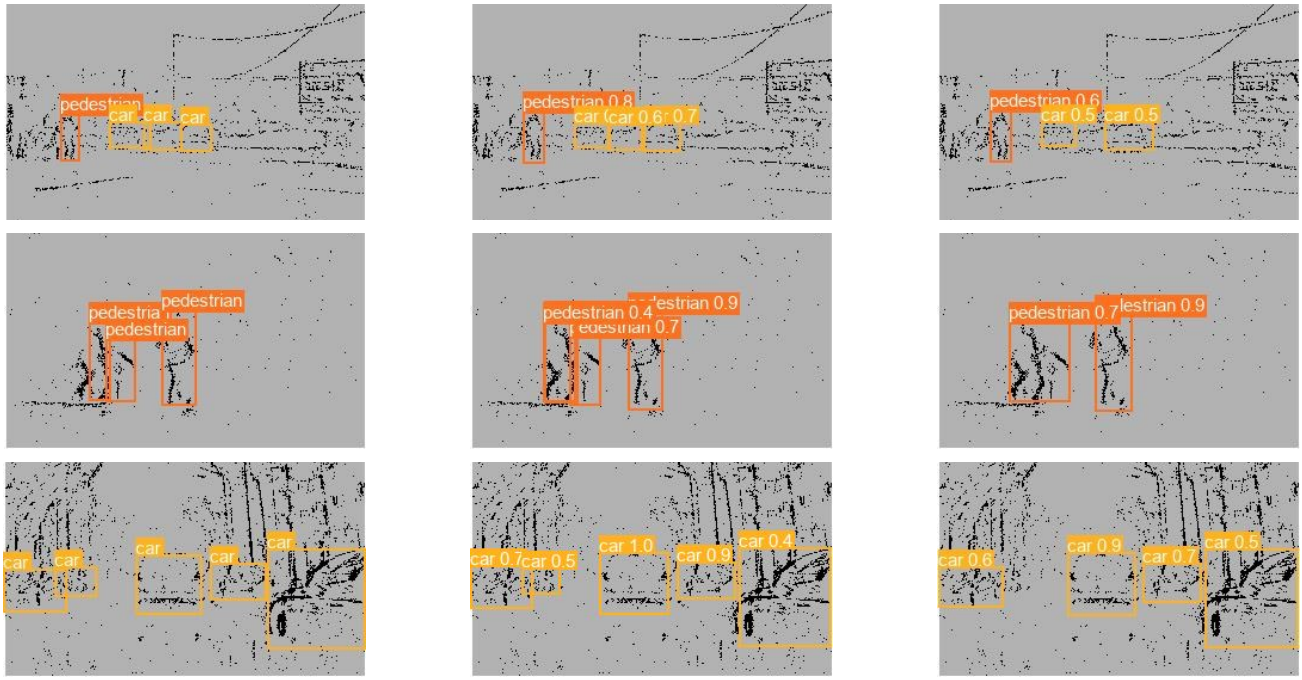

Figure 1. Visualizations of the object detection results on the Gen1 dataset. Compared with the Swin-Transformer v2, our proposed network (GET) detects most objects, while the bounding boxes are more accurate.

*Corresponding author

3. Video Object Detection Results

We provide a **video clip** in video.mp4. The video corresponds to an event clip in the validation set of the 1Mpx dataset. It includes the visualizations of ground truth and object detection results of GET and Swin-Transformer v2 [3]. In the video, our object detection results are significantly superior.

4. Head Details

For classification, the connected head consists of a global average pooling layer that aggregates the features from the last block of GET, followed by a fully connected layer that maps the features to the output classes.

For object detection, the connected head is the same as in RVT [2] (YOLOX [1] PAFPN and decoupled heads). The PAFPN depth is chosen as 0.67. The confidence and NMS thresholds are set as 0.001 and 0.45, respectively.

References

[1] Zheng Ge, Songtao Liu, Feng Wang, Zeming Li, and Jian Sun. Yolox: Exceeding yolo series in 2021. *arXiv preprint arXiv:2107.08430*, 2021. 2

[2] Mathias Gehrig and Davide Scaramuzza. Recurrent vision transformers for object detection with event cameras. *arXiv preprint arXiv: 2212.05598*, 2022. 2

[3] Ze Liu, Han Hu, Yutong Lin, Zhuliang Yao, Zhenda Xie, Yixuan Wei, Jia Ning, Yue Cao, Zheng Zhang, Li Dong, Furu Wei, and Baining Guo. Swin transformer v2: Scaling up capacity and resolution. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2