Supplementary Materials of HARDVS

Anonymous ICCV submission

Paper ID 3048

1. Dataset

In this work, three datasets are adopted for the evaluation of our proposed model, including **N-Caltech101** [2], **ASL-DVS** [1], and our newly proposed **HARDVS** as shown in Fig. 2. The distribution of each category in our HARDVS dataset is visualized in Fig. 1. More details about the **N-Caltech101** [2] and **ASL-DVS** [1] dataset are discussed below.

- **N-Caltech101 dataset** [2] [1] is derived from Caltech101 dataset which contains 100 different object categories and a background category. The number of samples in each category ranges from 40 to 800 and the resolution of each image is 300×200. This dataset was divided into the training set and the testing set at the ratio of 80% and 20%.
- **ASL-DVS dataset** [1] [2] is captured using iniLabs DAVIS240c camera and contains 24 categories with 4200 samples for each class. The resolution of each image is 240×180. This dataset was also divided into the training set and the testing set at the ratio of 80% and 20%.

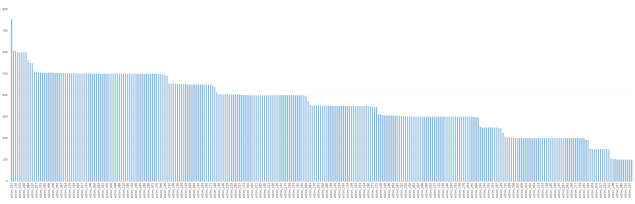

Figure 1. Number of videos in each class of our HARDVS dataset.

References

[1] Yin Bi, Aaron Chadha, Alhabib Abbas, Eirina Bourtsoulatze, and Yiannis Andreopoulos. Graph-based spatio-temporal feature learning for neuromorphic vision sensing. *IEEE Transactions on Image Processing*, 29:9084–9098, 2020.

[2] Garrick Orchard, Ajinkya Jayawant, Gregory K Cohen, and Nitish Thakor. Converting static image datasets to spiking neuromorphic datasets using saccades. *Frontiers in neuroscience*, 9:437, 2015.

[1] https://www.garrickorchard.com/datasets/n-caltech101

[2] https://www.dropbox.com/sh/ibq0jsicatn7l6r/AACNrNELV56rs1YInMWUs9CAa?dl=0

Figure 2. Visualization of 140 classes of HARDVS dataset.