# Asynchronous Events-based Panoptic Segmentation using Graph Mixer Neural Network

Sanket Kachole[1] Yusra Alkendi[2] Fariborz Baghaei Naeini[1,3] Dimitrios Makris[1] Yahya Zweiri[2]
Dept of Computer Science, Kingston University, London, UK[1] Ipsotek Ltd, London, UK [3]
Advanced Research and Innovation Center (ARIC), Khalifa University, Abu Dhabi, UAE[2]
{K1742163,f.baghaeinaeini,d.makris}@kingston.ac.uk[1] {yusra.alkendi,y.zweiri}@ku.ac.ae[2]

## 1. Dataset

The ESD dataset, as described in [3], is one of the largest datasets available for understanding instances of robotic grasping scenes. The dataset was captured using a Davis346 sensor mounted on a robotic arm and includes both conventional RGB frames and asynchronous events. The dataset also contains dense annotations for pixels and events, which are instance-specific and cover 15 classes that are grouped into 6 categories, namely bottle, box, pouch, book, mouse, and platform. The dataset comprises 17186 annotated images and 177 labeled event streams, with variations in the direction of camera motion, arm speed, lighting conditions, and object clutter. The motion variations include linear, rotational, and partial-rotational motion, while the arm speed variants are 0.15 m/s, 0.3 m/s, and 1 m/s. The lighting conditions comprise normal light and low light. The number of objects in the clutter varies from 2 to 10, as shown in figure 1. The training dataset includes 13984 images, the testing dataset includes 3202 images, and the validation dataset includes five objects that differ from those in the testing dataset.

## 2. Qualitative Evaluation

The qualitative results presented in 1, 2, 3 compares the performance of four different methods, namely EV-SegNet [2], ESS [4], GTNN [1], and GMNN (ours) for panoptic segmentation. The predictions in the table 1 were made on a dataset comprising eight objects, recorded in low light conditions with a rotational arm motion at a speed of 1 m/s, and the camera was positioned at a distance of 82 cm from the platform. The purpose of these experiments was to evaluate how effectively the objects are segmented to achieve the objective of panoptic segmentation.

The results of the study suggest that events in the dataset mainly overlap at the boundaries of occluded objects. The smaller objects are completely invisible in the EV-SegNet and ESS methods. While GTNN shows good results, GMNN performs even better. The object like a box that

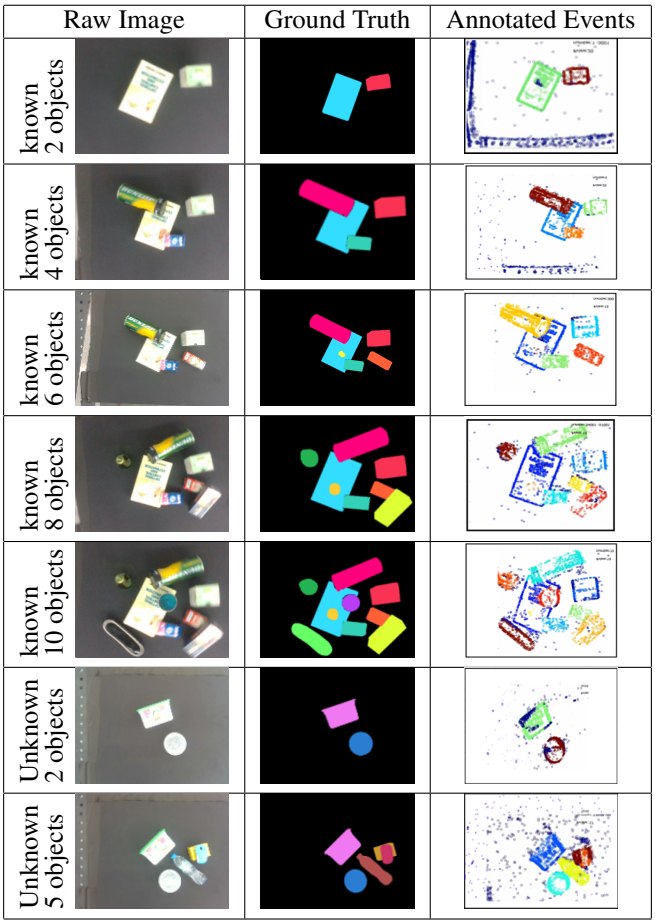


Table 1. Example of the ESD-1 dataset (row 1-5) in terms of the number of known objects attributes, under the condition of 0.15 moving speed, normal light condition, linear movement, and 0.82 height. The ESD-2 dataset (rows 6,7) presents examples of previously unseen objects with varying attributes. Specifically, the dataset features scenes where objects are moving at a speed of 0.15, under normal lighting conditions, with linear motion, and at a height of 0.82. The RGB ground truth and annotated event mask use different colors to represent different object labels. For optimal understanding, it is recommended to view the dataset in color.

has sharp corners was incorrectly segmented as background in all three methods except GMNN, which accurately segments such sharp edge objects. Additionally, the method proposed by the authors (GMNN) effectively handles occlusions. Interestingly, even in low light conditions, where fewer events were triggered, the proposed method (GMNN) achieved the highest quality. Overall, the results demonstrate that GMNN is a more effective method for panoptic segmentation, particularly in challenging scenarios where objects have sharp corners or are occluded, and under low light conditions where event triggering is reduced.

## References


[1] Yusra Alkendi, Rana Azzam, Sajid Javed, Lakmal Seneviratne, and Yahya Zweiri. Neuromorphic Vision-based Motion Segmentation with Graph Transformer Neural Network. Technical report. 1

[2] Iñigo Alonso and Ana C Murillo. EV-SegNet: Semantic Segmentation for Event-based Cameras. *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA*, pages 1624–1633, 2019. 1

[3] Xiaoqian Huang, Kachole Sanket, Abdulla Ayyad, Fariborz Baghaei Naeini, Dimitrios Makris, and Yahya Zweiri. A Neuromorphic Dataset for Object Segmentation in Indoor Cluttered Environment. *arXiv preprint arXiv:2302.06301*, 2 2023. 1

[4] Zhaoning Sun, Nico Messikommer, Daniel Gehrig, and Davide Scaramuzza. ESS: Learning Event-based Semantic Segmentation from Still Images. *arXiv preprint arXiv:2203.10016*, 2022. 1

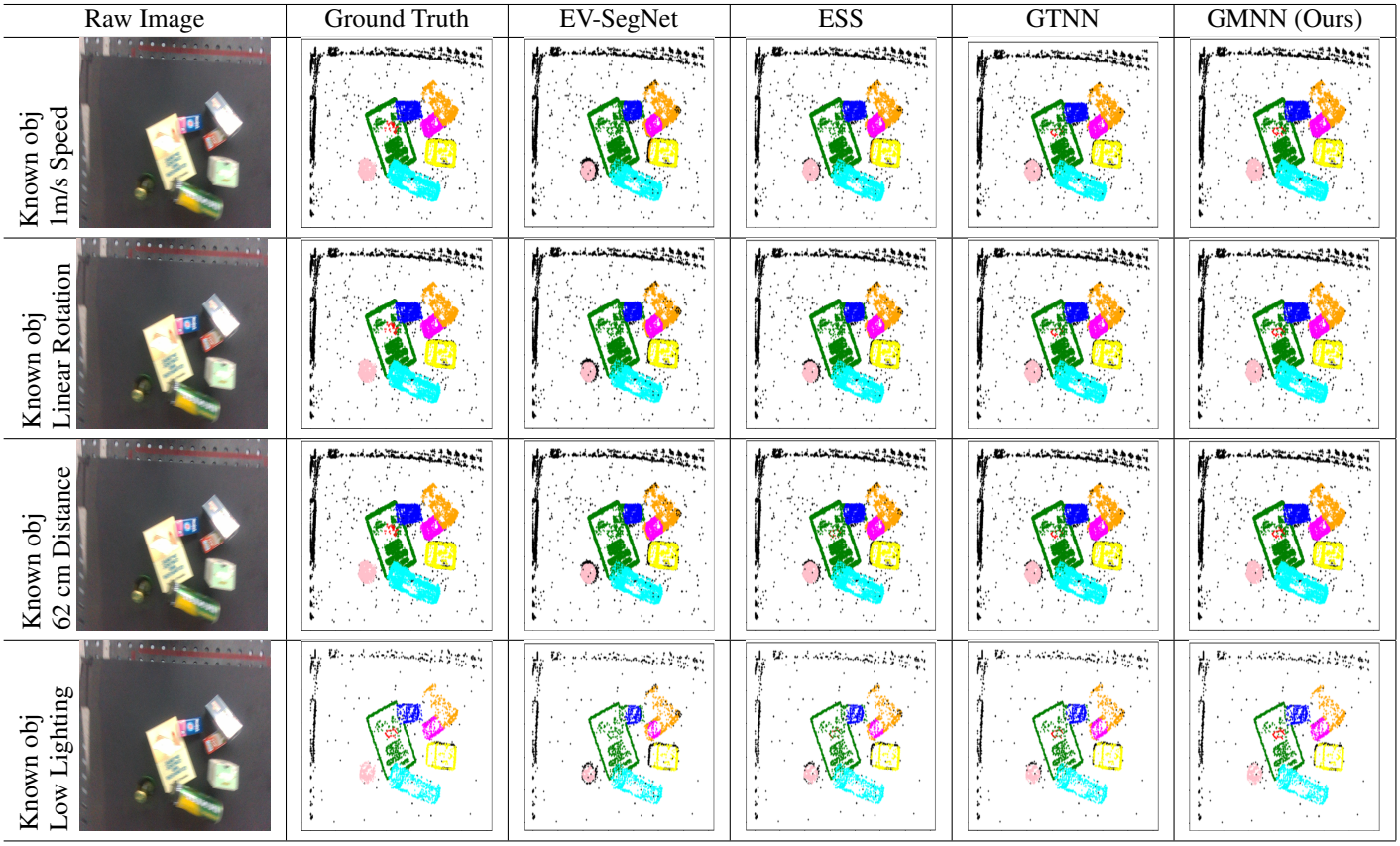


Figure 1. Qualitative Results - The qualitative results presented compares the performance of four different methods, namely EV-SegNet, ESS, GTNN, and GMNN (ours) for panoptic segmentation. The predictions were made on an ESD - 1 dataset i.e. known objects **displayed at row 1** comprising eight objects, recorded in bright light conditions with a rotational arm motion at a speed of **1 m/s**, and the camera was positioned at a distance of 82 cm from the platform. The predictions **displayed at row 2** comprising eight objects, recorded in bright light conditions with a **Linear** arm motion at a speed of 0.15 m/s, and the camera was positioned at a distance of 82 cm from the platform. The predictions **displayed at row 3** comprising eight objects, recorded in bright light conditions with a rotational arm motion at a speed of 0.15 m/s, and the camera was positioned at a distance of **62 cm** from the platform. The predictions **displayed at row 4** comprising eight objects, recorded in **Low light** conditions with a rotational arm motion at a speed of 0.15 m/s, and the camera was positioned at a distance of 82 cm from the platform.

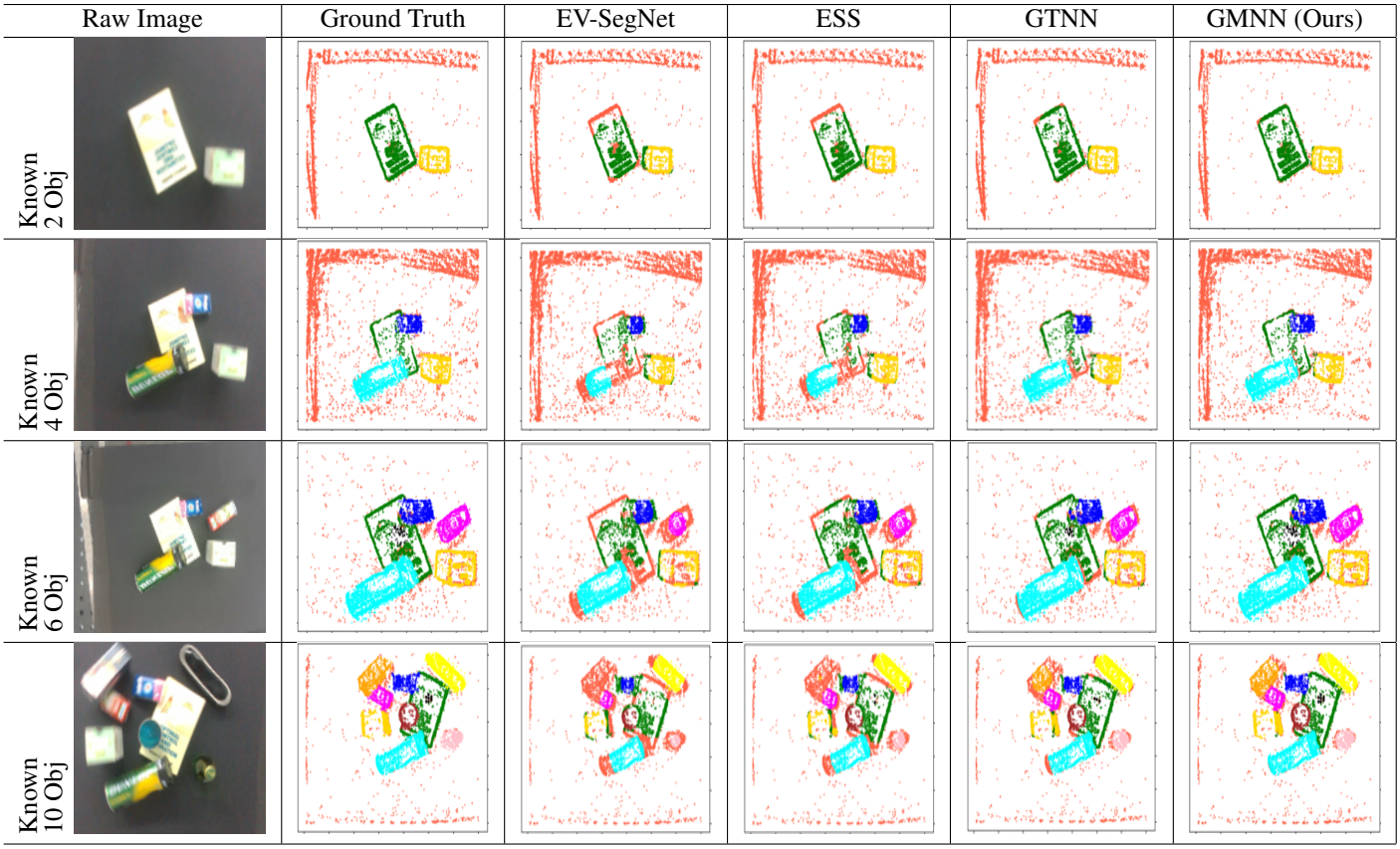


Figure 2. Qualitative Results - The qualitative results presented compares the performance of four different methods, namely EV-SegNet, ESS, GTNN, and GMNN (ours) for panoptic segmentation. The predictions were made on an **ESD-1 dataset i.e. known objects** comprising a varying number of objects in clutter, recorded in good light conditions with a rotational arm motion at a speed of 1 m/s, and the camera was positioned at a distance of 82 cm from the platform.

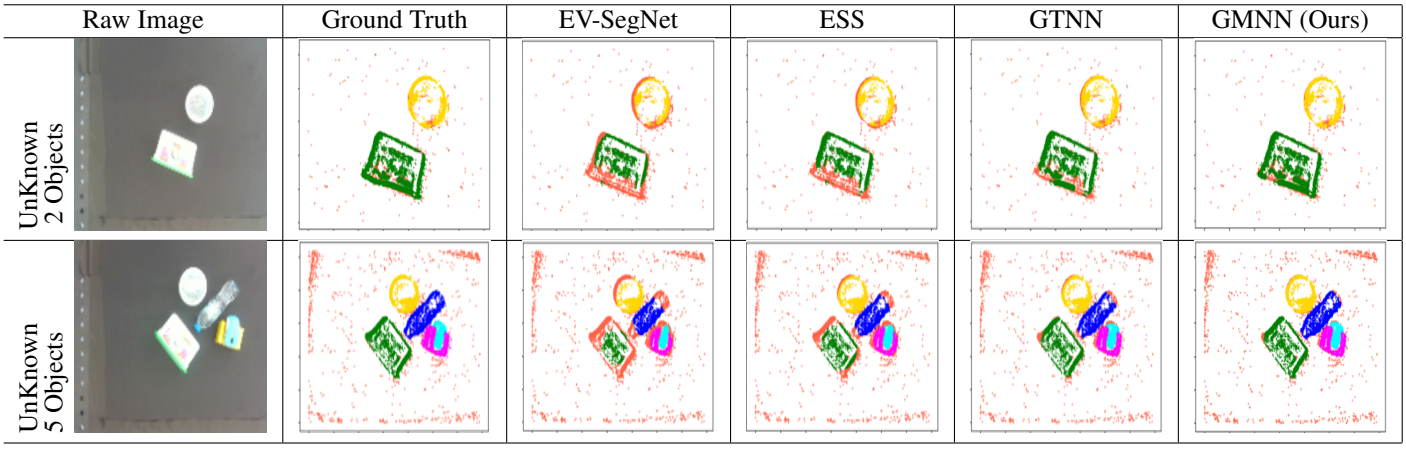


Figure 3. Qualitative Results - The qualitative results presented compares the performance of four different methods, namely EV-SegNet, ESS, GTNN, and GMNN (ours) for panoptic segmentation. The predictions were made on an **ESD-2 dataset i.e. Unknown objects** comprising a varying number of objects in clutter, recorded in good light conditions with a rotational arm motion at a speed of 1 m/s, and the camera was positioned at a distance of 82 cm from the platform.