Constructing and Evaluating Complex Event-based Datasets for Increasing Performance of Instance Segmentation Models

1. Background

Event-based cameras output a stream of events for each pixel.

They offer **advantages**:

- low output bandwidth
- low latency
- high dynamic range

Segmentation is a computer vision task, where each pixel is labeled.

Consider constructing event-based datasets with noisy background, by a **superimposition** of two event datasets, and by adding unifrom noise.

2. Question

Would **noisy datasets** bring improvements over the original event datasets for simulating noisy, real-world environments thereby **increasing** the performance of **segmentation Machine Learning models**?

- *What is the optimal amount of noise that can be superimposed from an existing event-based dataset over a different event-based dataset to gain an improvement in segmentation tasks?*
- *How does applying random noise over an existing event-based dataset affect the performance of instance segmentation models?*

3. Method

Constructing noisy ***event-based datasets*** *and evaluating them on an* ***instance segmentation*** *model.*

1. Two datasets superimposed: N-MNIST and N-Caltech101 [1].

2. Three resulting datasets, with increasing amounts of noise:
   - **Superimposed-Noisy**
   - **Centered-Filtered**
   - **Uniform-Noisy**

3. Instance segmentation model used: **EV-Mask-RCNN** [2].

4. Three instances of the model are trained on each of the four datasets, resulting a total of 12 trained models.

5. Best model per dataset is evaluated on the original N-MNIST test set multiple times (**cross evaluations**).

6. Perform two-tailed, independent T-Test of all recorded metrics to determine **if values are statistically significant**.

4. Results

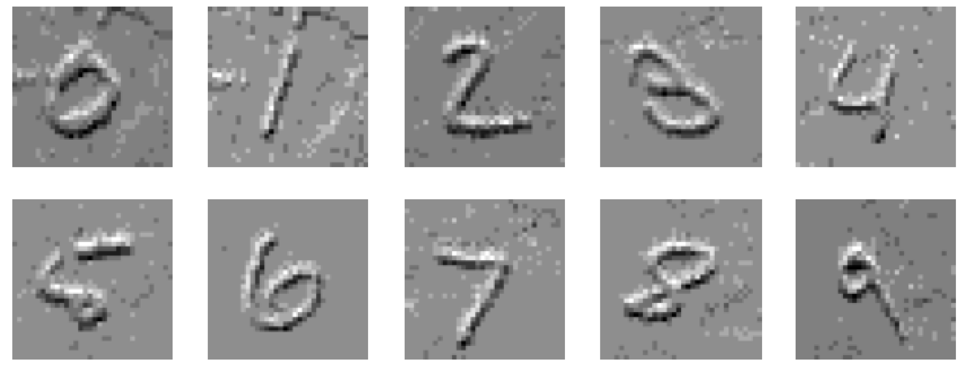

Figure 1. Sample of each digit from ***Centered-Filtered*** *dataset .*

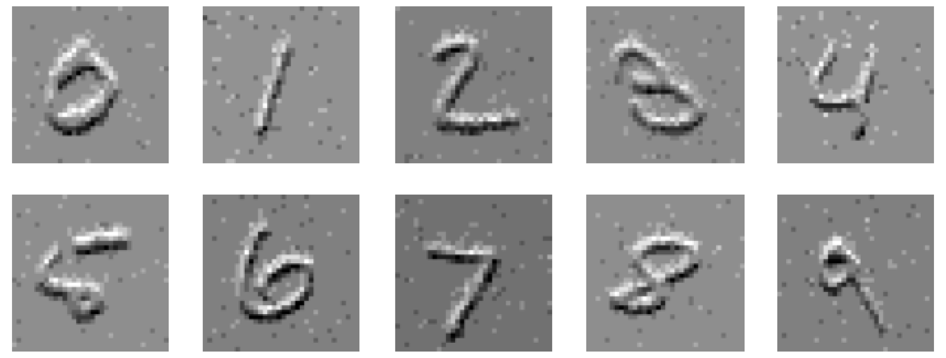

Figure 2. Sample of each digit from ***Uniform-Noisy*** *dataset.*

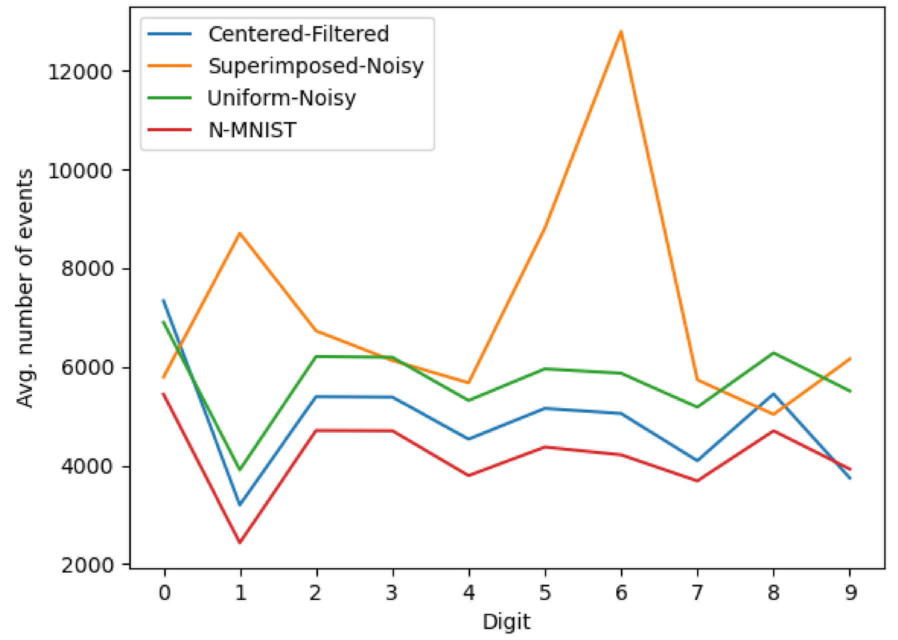

Figure 3. Average amount of noise per digit for each generated dataset.

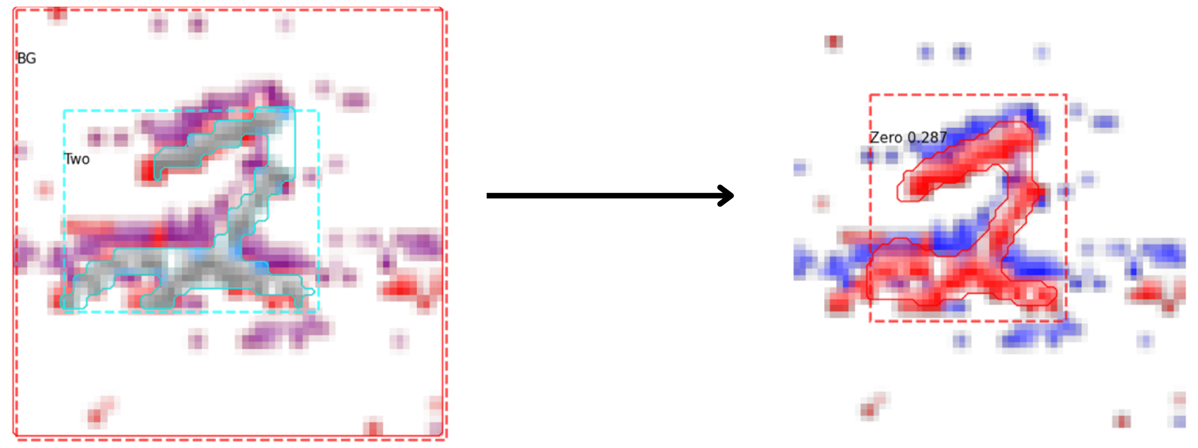

Figure 4. (A) Ground truth of digit 2.
(B) Segmentation mask and class prediction.

Model	mAcc	mAP	mIoU
Superimposed-Noisy	94.75	21.97	24.36
Centered-Filtered	95.12	28.44	34.24
Uniform-Noisy	95.31	29.3	36.58
No-noise	95.53	33.71	42.3

Figure 5. Results of training on noisy datasets and testing on N-MNIST test set .

5. Conlcusion

1. **Results of T-Test** p-values are **smaller** than *0.001*, indicating no relationships between the sets of data.

2. Models trained on noisy datasets perform **noticeably worse** on a dataset with no noise, than a model trained directly on the no-noise dataset.

3. **Uniform-Noisy** has more events than Centered-Filtered, but performed best in cross evaluations - uniformity of noise allowed it to generalize better.

4. Altering the **background** makes predictions generalizabiliy decrease. Altering the **data (digits)** => increase [3].

Future work: increase the variability of noise between the datasets, vary time-windows of training data, evaluate using other models

References:

[1] G. Orchard, A. Jayawant, G. K. Cohen, and N. Thakor, "Converting static image datasets to spiking neuromorphic datasets using saccades," Frontiers in neuroscience, vol. 9, p. 437, 2015.

[2] A. Baltaretu, "EV-Mask-RCNN: Instance segmentation in event-based videos," 2022.

[3] F. Gu, W. Sng, X. Hu, and F. Yu, "Eventdrop: data augmentation for event-based learning," arXiv preprint arXiv:2106.05836, 2021

Alexandru-Dragos Manolache – a.d.manolache@student.tudelft.nl | Supervised by: Nergis Tömen, Ombretta Strafforello, Xin Liu

CSE3000 – Research Project | June 22, 2022