Bridging the Gap Between RGB & Event Cameras

Aninda Ghosh

What?

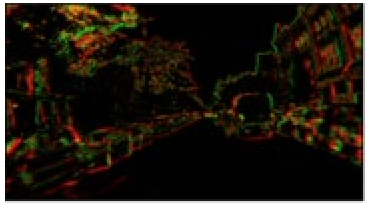
Our Goal

What problem are we trying to address?

- Event Cameras are new, but the concepts aren't.
- Hence, an event-based vision has been held back by the shortage of labeled datasets.
- We propose a task transfer method to train models directly with labeled images and unlabeled event data.

Real Image

Label



Event Data

???

Hi! I am an event camera.

Why?

Event Cameras over Lidar and RGB Cameras

Perspectives from Utility, Efficiency, & Economy

Higher Temporal Resolution

- High temporal resolution μs range

- Can detect and track fast-moving objects in real-time

Energy Efficiency

Consumes very little power compared to LiDAR

High Dynamic Range

Has a high dynamic range, which means they can capture images in both bright and dark environments.

Low Latency

Has low latency, which means they can provide real-time feedback to the control system.

No Moving Parts

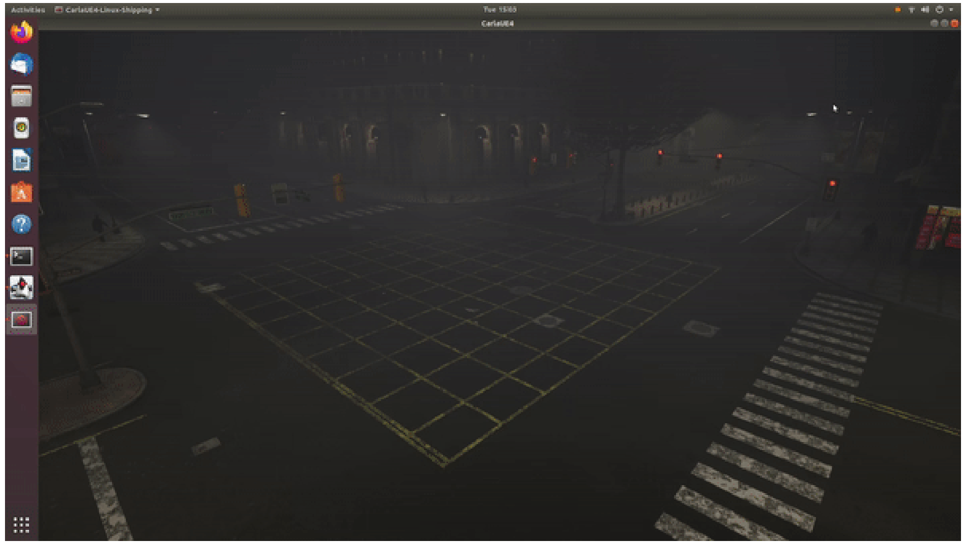
Do not have any moving parts, making them more reliable and less prone to mechanical failures.

Robustness

Resistant to motion blur, vibration and performs well even in rain/snow.

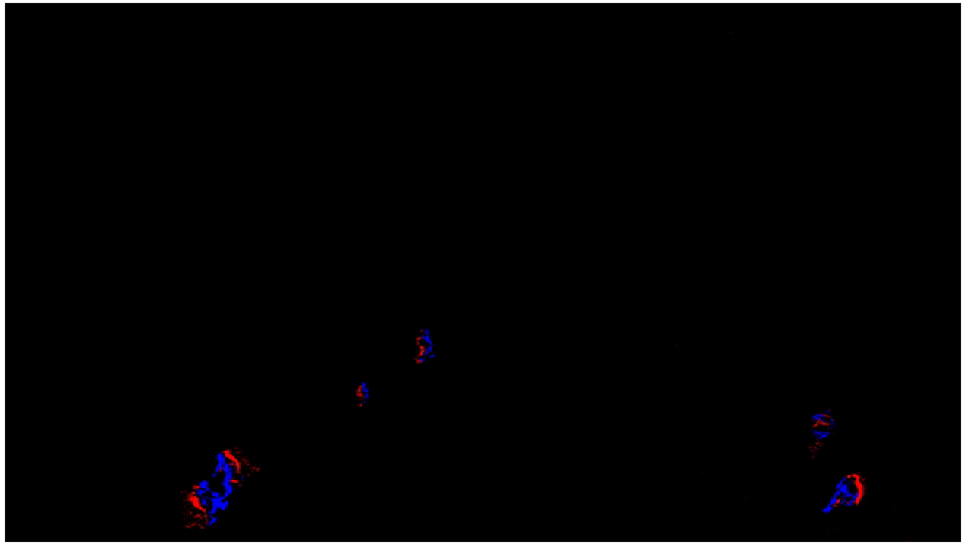
Example Use Cases

Capturing Motion in Dark & Foggy Environments



RGB Camera capture in Foggy and Night Weather

Event Camera capture at night

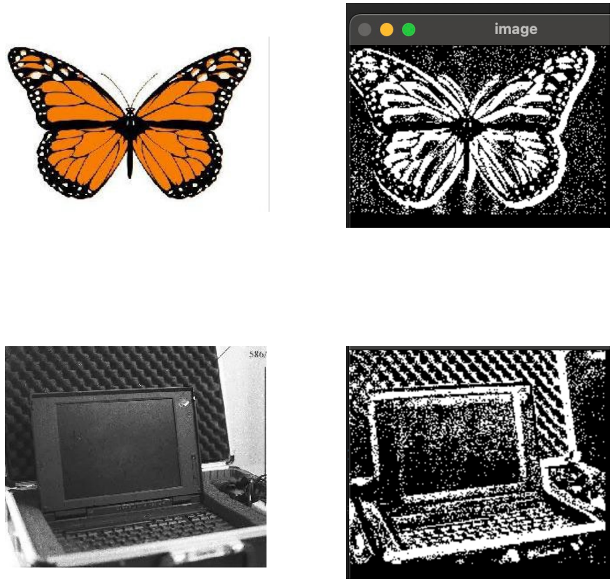


Event Camera capture in Night+Foggy Weather

Inspiration

Nico Messikommer, Daniel Gehrig, Mathias Gehrig, Davide Scaramuzza. Bridging the Gap between Events and Frames through Unsupervised Domain Adaptation.

N-Caltech 101 Dataset



Bridging the Gap between Events and Frames through Unsupervised Domain Adaptation

Nico Messikommer, Daniel Gehrig, Mathias Gehrig, Davide Scaramuzza

Abstract—**Reliable perception during fast motion maneuvers or in high dynamic range environments is crucial for robotic systems. Since event cameras are robust to these challenging conditions, they have great potential to increase the reliability of robot vision. However, event-based vision has been held back by the shortage of labeled datasets due to the novelty of event cameras. To overcome this drawback, we propose a task transfer method to train models directly with labeled images and unlabeled event data. Compared to previous approaches, (i) our method transfers from single images to events instead of high frame rate videos, and (ii) does not rely on paired sensor data. To achieve this, we leverage the generative event model to split event features into content and motion features. This split enables efficient matching between latent spaces for events and images, which is crucial for successful task transfer. Thus, our approach unlocks the vast amount of existing image datasets for the training of event-based neural networks. Our task transfer method consistently outperforms methods targeting Unsupervised Domain Adaptation for object detection by 0.26 mAP (increase by 93%) and classification by 2.7% accuracy.**

Index Terms—**Deep Learning for Visual Perception, Object Detection, Segmentation and Categorization, Transfer Learning**

MULTIMEDIA MATERIAL

The code of this project is available at https://github.com/

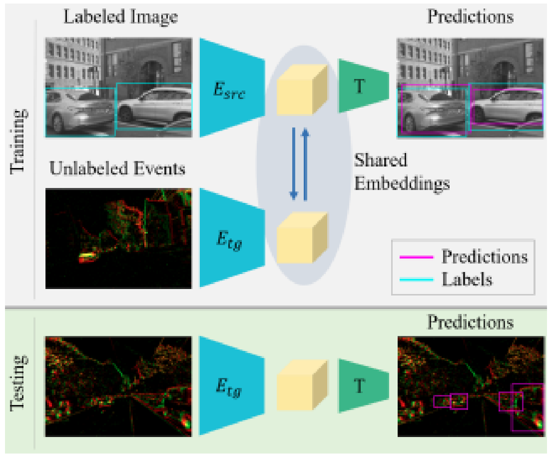


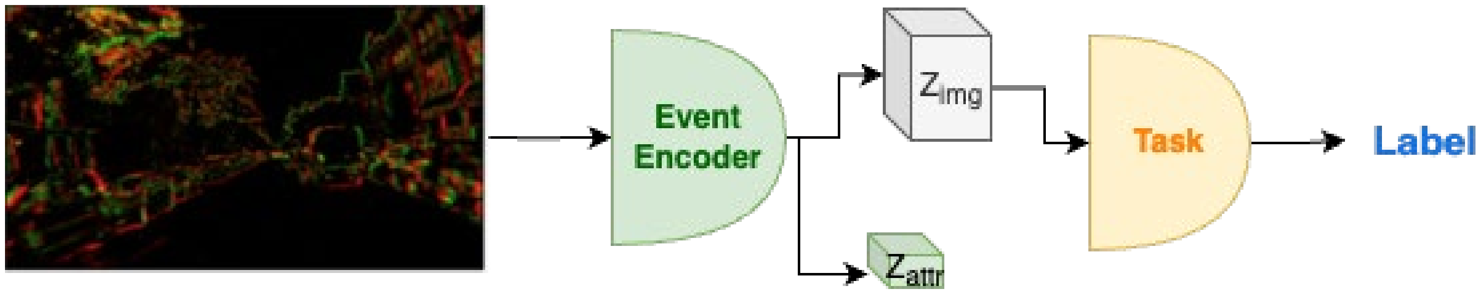
Fig. 1: Our approach can teach a network to detect cars in event frames even though it was never told how cars look in the event domain. This unsupervised domain adaption is possible by leveraging labeled grayscale images and unlabeled events. During testing, our approach consists of a simple encoder E and task network T and thus has no computational overhead of first translating events to images.

How?

Baseline

Architecture

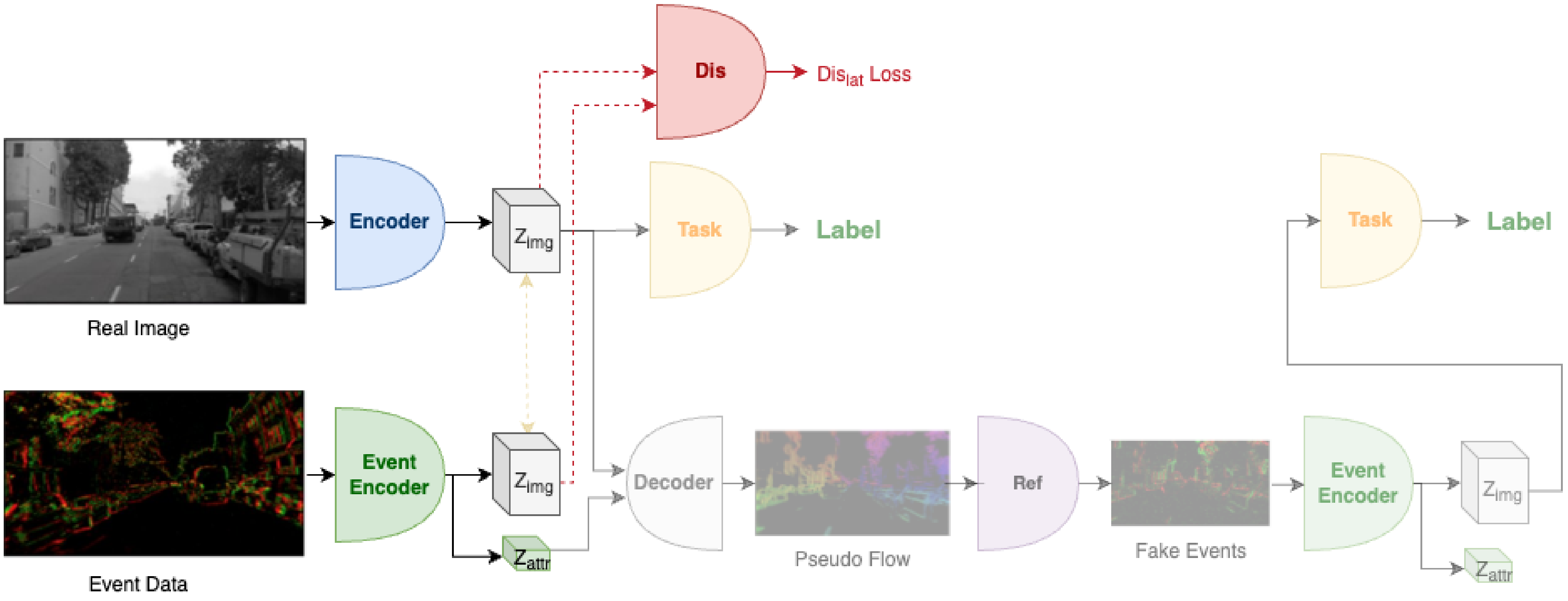
Inference Model



Baseline

Architecture

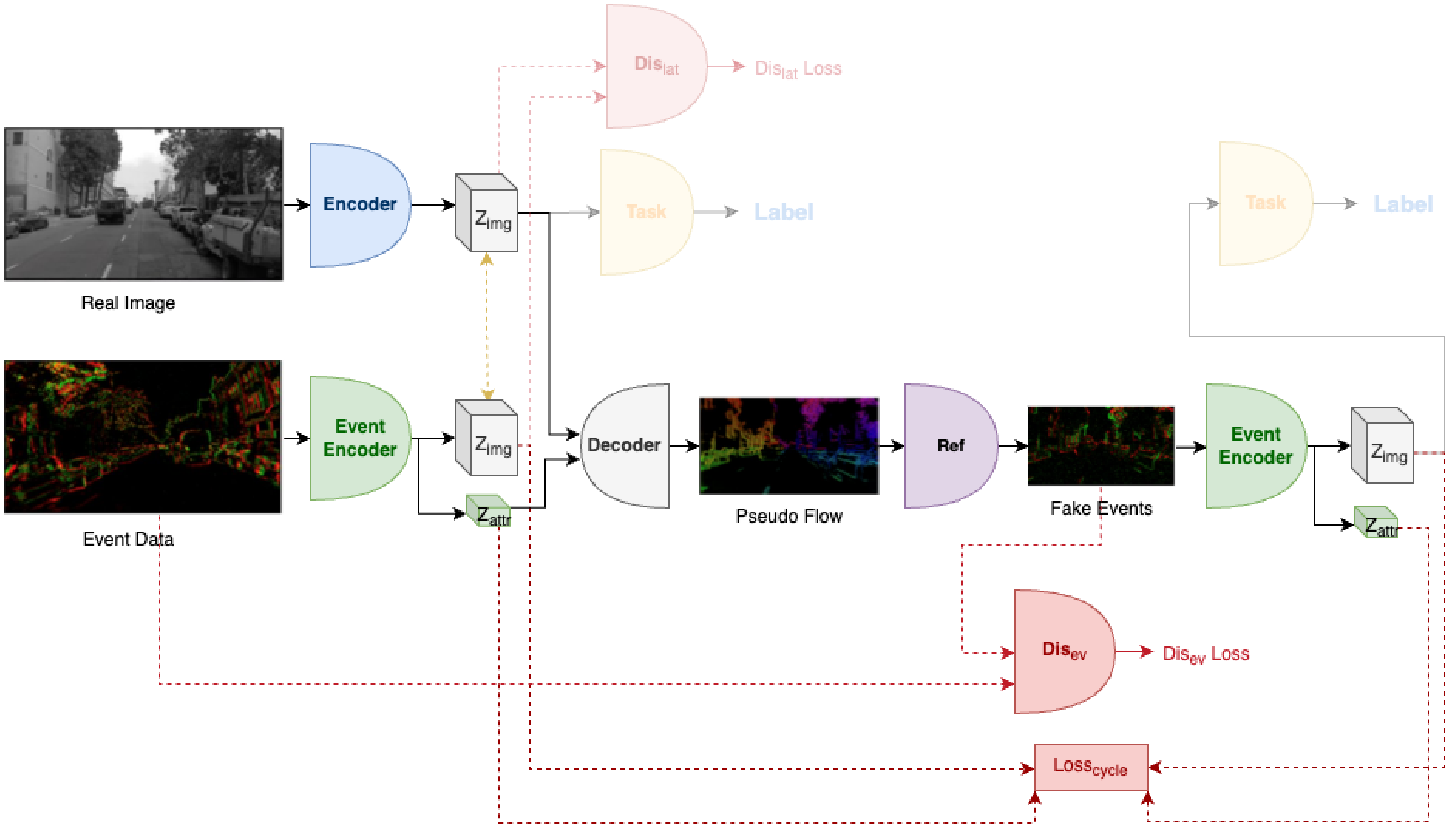
Align spatial embedding using GAN based training



Baseline

Architecture

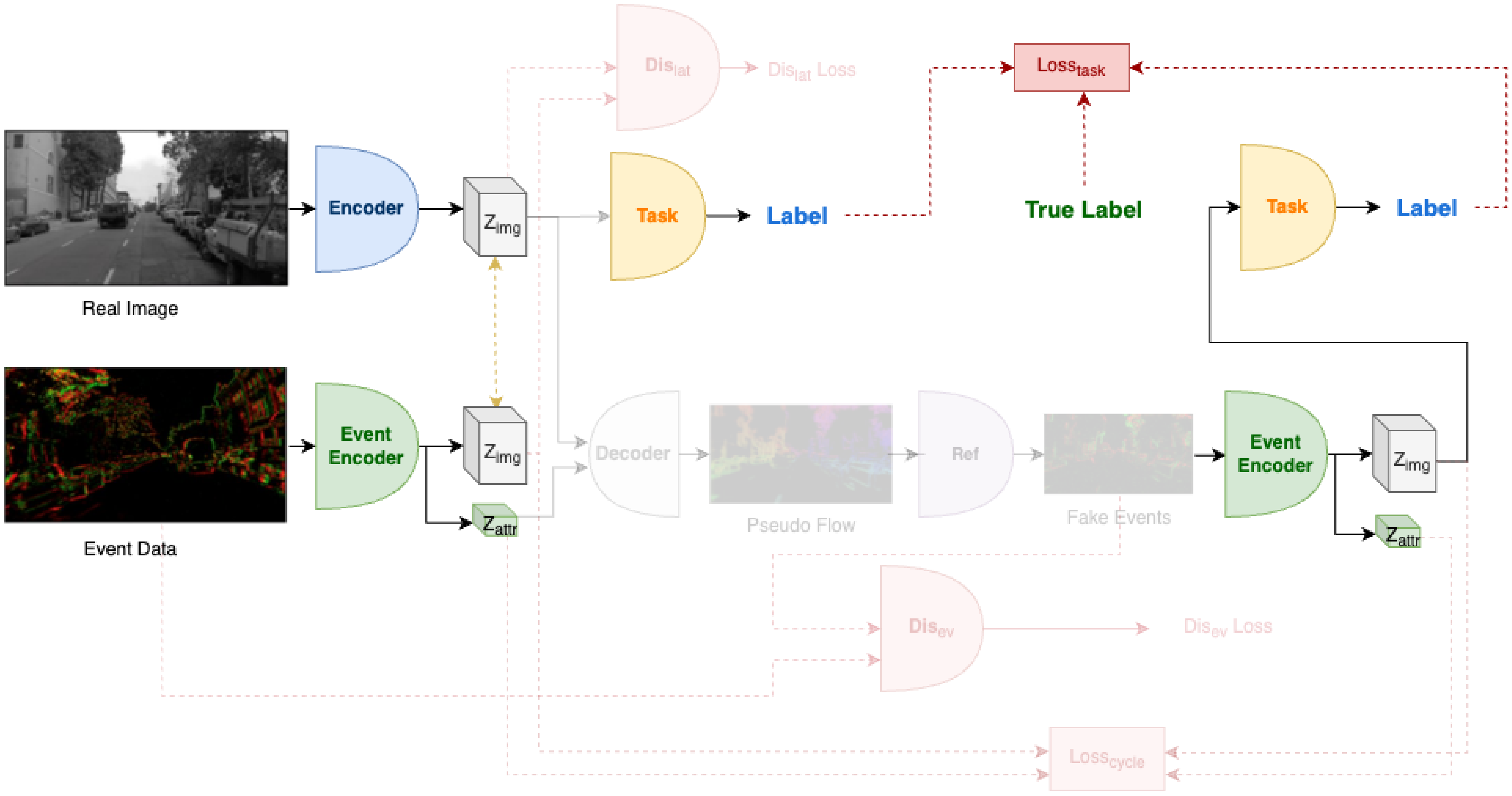
Generating Fake events and aligning the event embeddings



Baseline

Architecture

Learning the task



Can There be Improvements?

Changes We Did

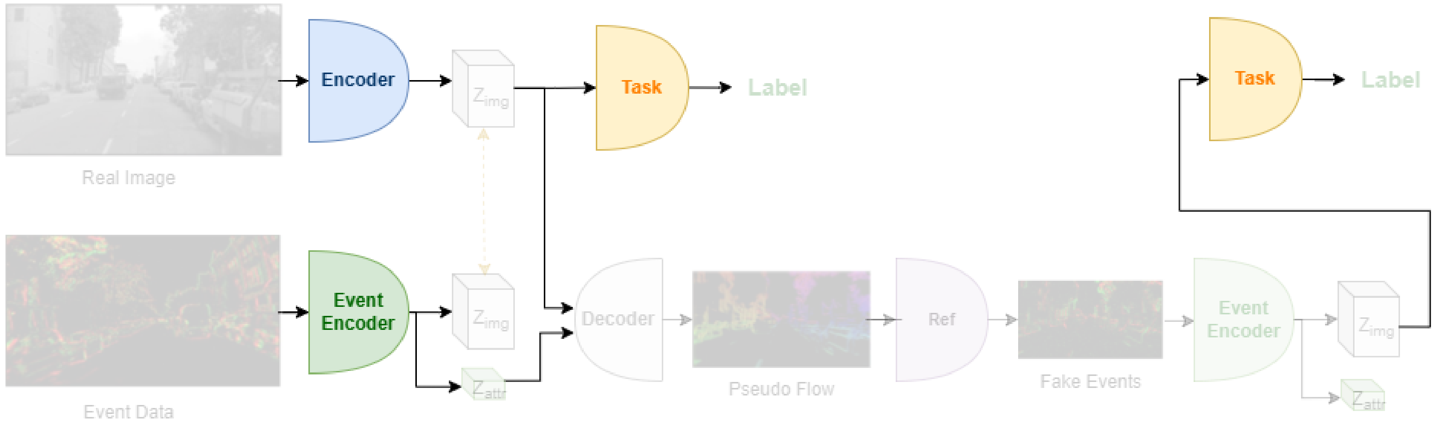
1. Changing from Basic Resnet blocks to Bottleneck Resnet Blocks (With Sliced Input Dimension)

- Layer 1
  - Basic Block 1
  - Basic Block 2
- Layer 2
  - Basic Block 1
  - Basic Block 2

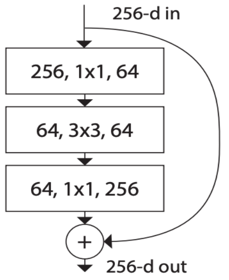
- Layer 1
  - Bottleneck Block 1
  - Bottleneck Block 2
  - Bottleneck Block 3
  - Bottleneck Block 4
- Layer 2
  - Bottleneck Block 1
  - Bottleneck Block 2

- Layer 3
  - Basic Block 1
  - Basic Block 2
- Layer 4
  - Basic Block 1
  - Basic Block 2

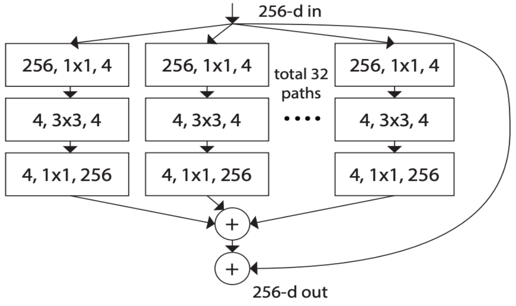
- Layer 2
  - Bottleneck Block 3
  - Bottleneck Block 4
- Layer 3
  - Bottleneck Block 1
  - Bottleneck Block 2
  - Bottleneck Block 3
  - Bottleneck Block 4
- Layer 4
  - Bottleneck Block 1
  - Bottleneck Block 2
  - Bottleneck Block 3
  - Bottleneck Block 4



Basic Blocks



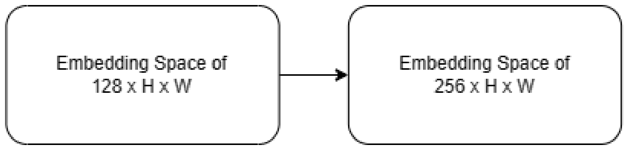
Parallelized Bottleneck Blocks

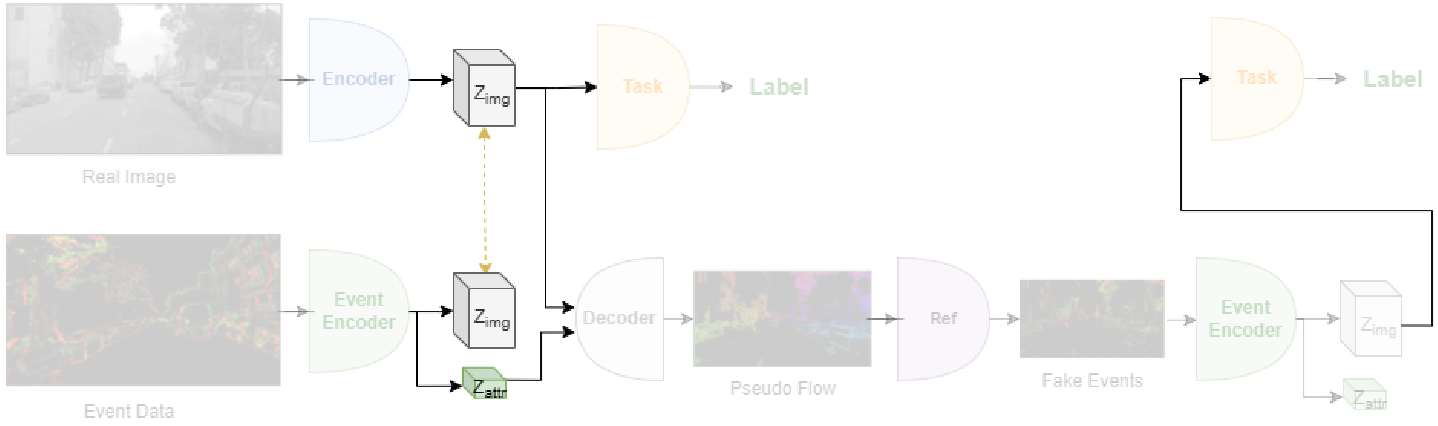


Improved Training Speed!

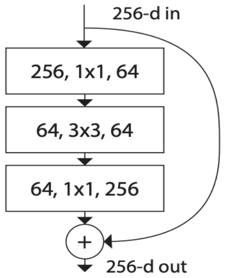
Changes We Did

2. Increasing the embedding space without compromising the performance.

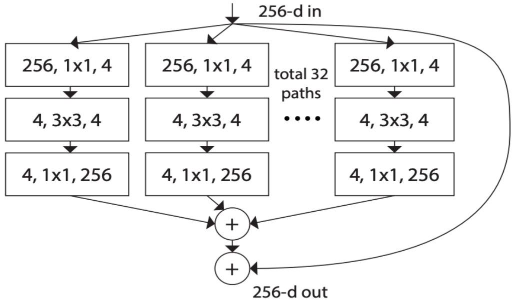




Basic Blocks



Parallelized Bottleneck Blocks



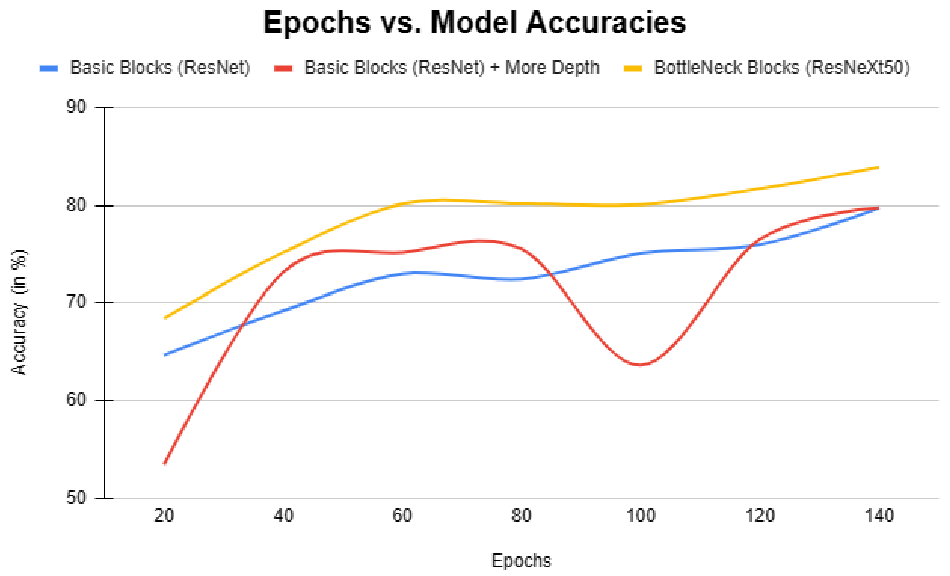
Improved Accuracy!

Results!

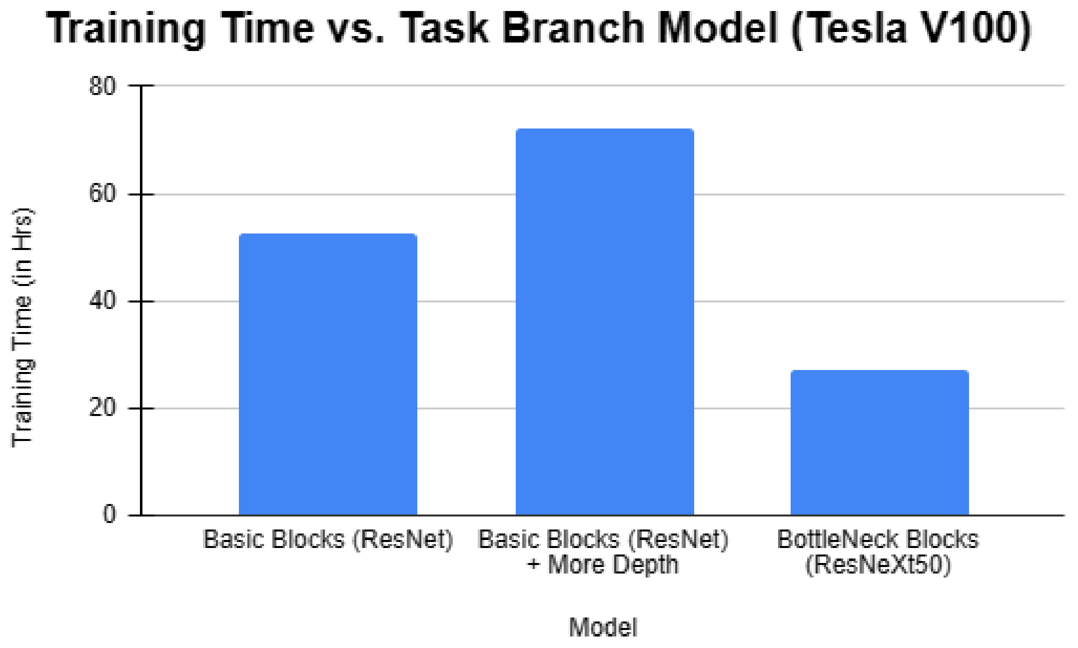
Metric

Measuring in terms of Accuracy and Training Time

Epochs	Basic Blocks (ResNet)	Basic Blocks (ResNet) + More Depth	BottleNeck Blocks (ResNeXt50)
20	64.63	53.41	68.39
40	69.19	73.15	75.15
60	72.95	75.15	80.11
80	72.4	75.5	80.16
100	75.06	63.58	80.06
120	75.96	76.5	81.66
140	79.71	79.75	83.87



Model	Training Time (hrs)
Basic Blocks (ResNet)	52.5
Basic Blocks (ResNet) + More Depth	72
BottleNeck Blocks (ResNeXt50)	27



References

- A. Dosovitskiy, G. Ros, F. Codevilla, A. Lopez, and V. Koltun, “Carla: An open urban driving simulator,” in Conference on robot learning, pp. 1–16, PMLR, 2017.
- N. Messikommer, D. Gehrig, M. Gehrig, and D. Scaramuzza, “Bridging the gap between events and frames through unsupervised domain adaptation,” IEEE Robotics and Automation Letters, vol. 7, no. 2, pp. 3515–3522, 2022.
- P. De Tournemire, D. Nitti, E. Perot, D. Migliore, and A. Sironi, “A large scale event-based detection dataset for automotive,” arXiv preprint arXiv:2001.08499, 2020.
- A. Tomy, A. Paigwar, K. S. Mann, A. Renzaglia, and C. Laugier, “Fusing event-based and rgb camera for robust object detection in adverse conditions,” in 2022 International Conference on Robotics and Automation (ICRA), pp. 933–939, IEEE, 2022.
- Y. Zhao, X. Liang, X. Fan, Y. Wang, M. Yang, and F. Zhou, “Mvsec: multi-perspective and deductive visual analytics on heterogeneous network security data,” Journal of Visualization, vol. 17, pp. 181–196, 2014.
- J. Binas, D. Neil, S.-C. Liu, and T. Delbruck, “Ddd17: End-to-end davis driving dataset,” arXiv preprint arXiv:1711.01458, 2017.
- P. Sun, H. Kretzschmar, X. Dotiwalla, A. Chouard, V. Patnaik, P. Tsui, J. Guo, Y. Zhou, Y. Chai, B. Caine, et al., “Scalability in perception for autonomous driving: Waymo open dataset,” in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 2446–2454, 2020.
- E. Perot, P. De Tournemire, D. Nitti, J. Masci, and A. Sironi, “Learning to detect objects with a 1-megapixel event camera,” Advances in Neural Information Processing Systems, vol. 33, pp. 16639–16652, 2020.
- H. Rebecq, D. Gehrig, and D. Scaramuzza, “Esim: an open event camera simulator,” in Conference on robot learning, pp. 969–982, PMLR, 2018.
- P. R. G. Cadena, Y. Qian, C. Wang, and M. Yang, “Spade-e2vid: Spatially-adaptive denormalization for event-based video reconstruction,” IEEE Transactions on Image Processing, vol. 30, pp. 2488–2500, 2021.
- J. Hidalgo-Carrió, D. Gehrig, and D. Scaramuzza, “Learning monocular dense depth from events,” in 2020 International Conference on 3D Vision (3DV), pp. 534–542, IEEE, 2020.
- C. Reinbacher, G. Graber, and T. Pock, “Real-time intensity-image reconstruction for event cameras using manifold regularization,” arXiv preprint arXiv:1607.06283, 2016
- https://www.garrickorchard.com/datasets/n-caltech101

Questions? & Feedbacks!