Paper ID: 6735

SFOD: Spiking Fusion Object Detector

Yimeng Fan[1], Wei Zhang[1,2,*], Changsong Liu[1], Mingyang Li[1], Wenrui Lu[1]

[1]School of Microelectronics, Tianjin University, China
[2]Tianjin Key Laboratory of Low-dimensional Electronic Materials and Advanced Instrumentation

Introduction

Motivation: Why do we need Spiking Fusion in SNNs?

- **The combination of deeper and shallower feature maps in the spatial domain**
- **Enhancing connections between features of different scales in the temporal domain**

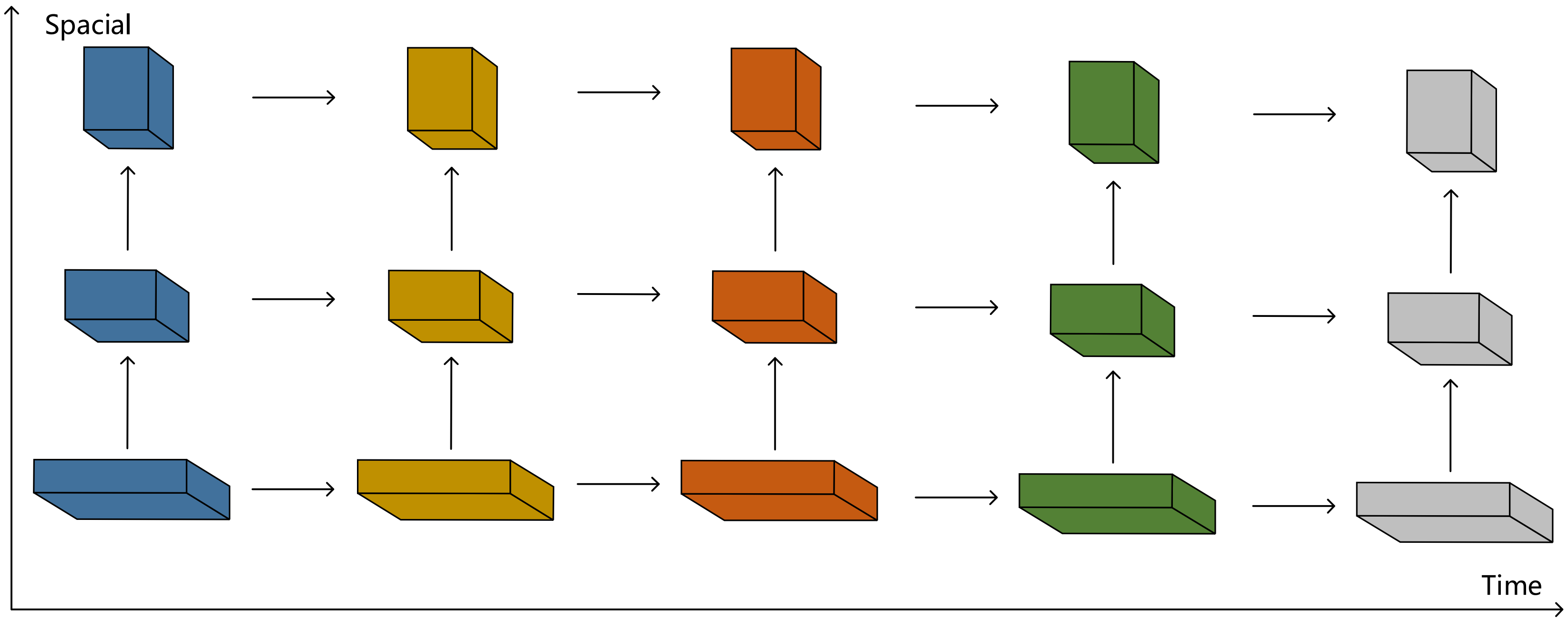

Fig1. Spatiotemporal feature extraction

Introduction

Related Work: Lack of corresponding research on Spiking Fusion

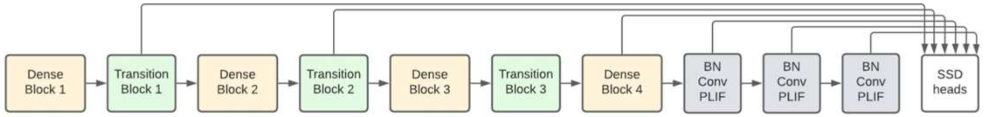

Fig2. Overview of Spiking DenseNet + SSD architecture [1].

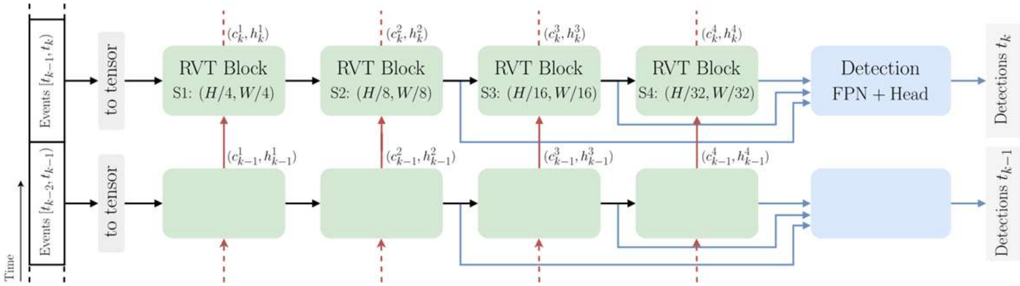

Fig3. Overview of RVT model [2].

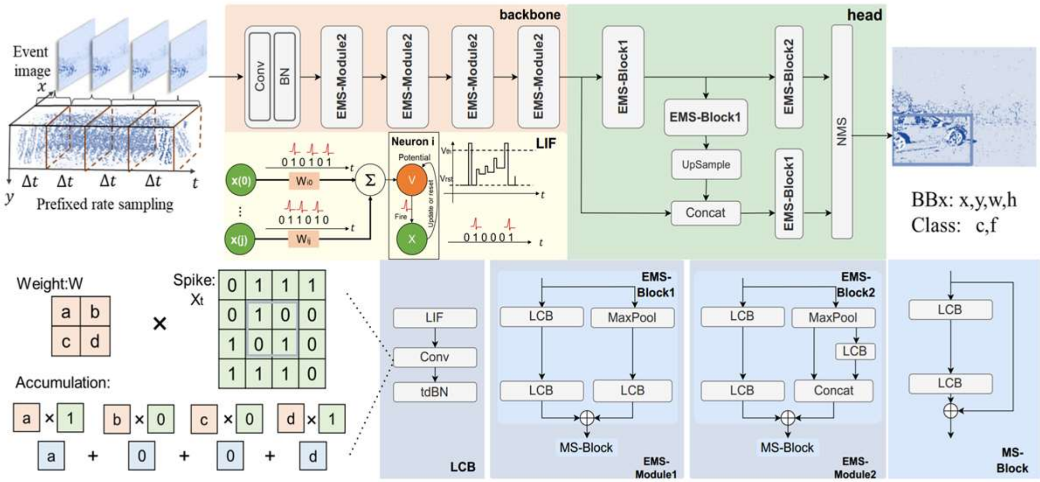

Fig4. Overview of EMS-YOLO model [3].

[1] Cordone L, Miramond B, Thierion P. Object detection with spiking neural networks on automotive event data[C]//2022 International Joint Conference on Neural Networks (IJCNN). IEEE, 2022: 1-8.
[2] Gehrig M, Scaramuzza D. Recurrent vision transformers for object detection with event cameras[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 13884-13893.
[3] Su Q, Chou Y, Hu Y, et al. Deep directly-trained spiking neural networks for object detection[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023: 6555-6565.

Method

Simple Fusion Model: SFOD(Spiking Fusion Object Detector)

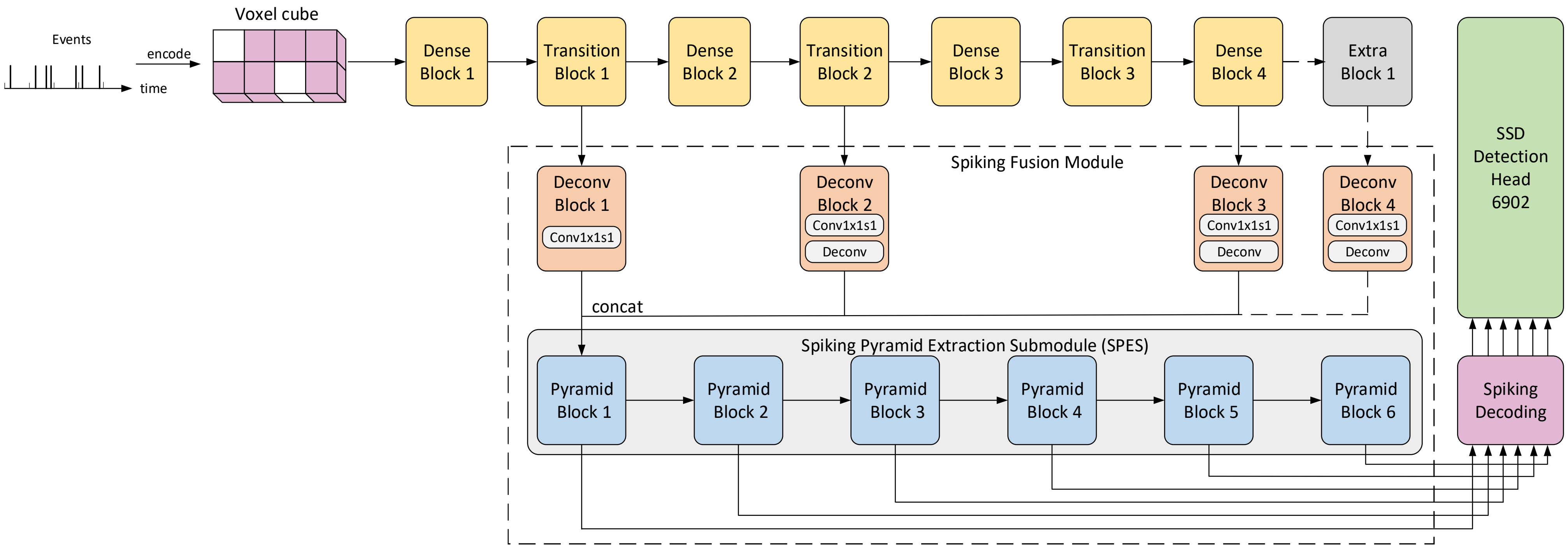

Fig5. Overview of SFOD.

Method

The architectures of SPES

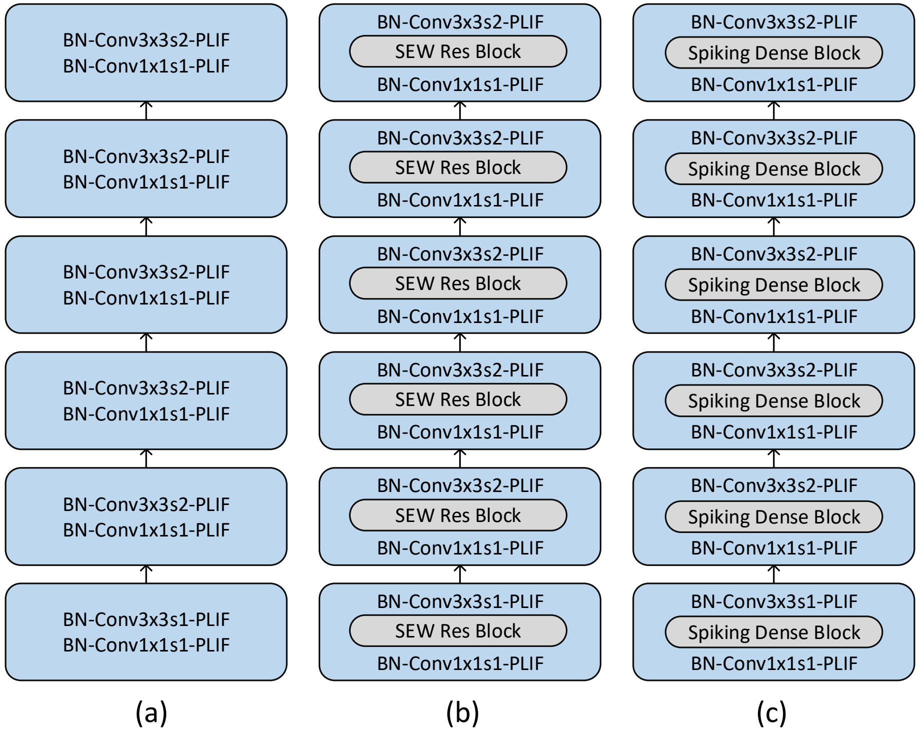

Fig6. The architecture of SPES.

Experiments

CVPR SEATTLE, WA JUNE 17-21, 2024 天津大学 Tianjin University

Ablation Results:

Models	Dec.		Fusion Layers			Params	mAP@0.5:0.95	mAP@0.5	Firing Rate
	Rate	Count	None	3	4				
DenseNet121-16-SSD	✓		✓			5.0M	0.262	0.517	21.01%
DenseNet121_24-SSD		✓	✓			8.2M	0.235	0.445	22.02%
DenseNet121-24-SSD	✓		✓			8.2M	0.288	0.553	22.29%
DenseNet169-16-SSD	✓		✓			7.7M	0.257	0.507	22.82%
SFOD-B	✓				✓	15.0M	0.294	0.570	21.13%
SFOD-B	✓			✓		9.9M	0.299	0.575	24.41%
SFOD-D	✓			✓		11.3M	0.286	0.558	26.37%
SFOD-R	✓			✓		**11.9M**	**0.321**	**0.593**	**24.04%**

Tab1. Results of the ablation study on the GEN1 dataset. We first study the performance differences across object detection models using various backbone networks. Based on this, we select the best backbone and further analyze the impact of different fusion layers. Finally, we compare the performance of various SPES variants. We name the models using the basic, Spiking Dense Block-enhanced, and SEW Res Block-enhanced SPESs as SFOD-B, SFOD-D, and SFOD-R, respectively.

Experiments

Benchmark Comparisons:

Method	Networks	Detection Head	Params	mAP @0.5:0.95	Firing Rate	Time (ms)	Energy (mJ)
Asynet [30]	Sparse CNNs	YOLOv1 [34]	11.4M	0.145	-	-	> 4.83
AEGNN [39]	GNNs	YOLOv1	20.0M	0.163	-	-	-
Inception+SSD [17]	CNNs	SSD [25]	-	0.301	-	19.4	-
MatrixLSTM [5]	RNNs+CNNs	YOLOv3 [33]	61.5M	0.310	-	-	-
RED [32]	RNNs+CNNs	SSD	24.1M	0.400	-	16.7	> 24.08
RVT [15]	**Transformer+RNNs**	**YOLOX [14]**	**18.5M**	**0.472**	**-**	**10.2**	**-**
MobileNet-64+SSD [9]	SNNs	SSD	24.3M	0.147	29.44%	$1.7^{\dagger}$	5.76
VGG-11+SDD [9]	SNNs	SSD	12.6M	0.174	22.22%	$4.4^{\dagger}$	11.06
DenseNet121-24+SSD [9]	SNNs	SSD	8.2M	0.189	37.20%	$4.1^{\dagger}$	3.89
EMS-YOLO [41]	**SNNs**	**YOLOv3**	**14.4M**	**0.310**	**17.80%**	**-**	**-**
SFOD	**SNNs**	**SSD**	**11.9M**	**0.321**	**24.04%**	**6.7**	**7.26**

Tab2. Comparison with state-of-the-art models on the GEN1 dataset. We present a comparison of our model with other state-of-the-art approaches on the GEN1 dataset. Remarkably, our model achieves a state-of-the-art mAP of 32.1% at the same level of firing rate and parameters compared to other SNN-based methods.

Inference results of the model on the GEN1 dataset:

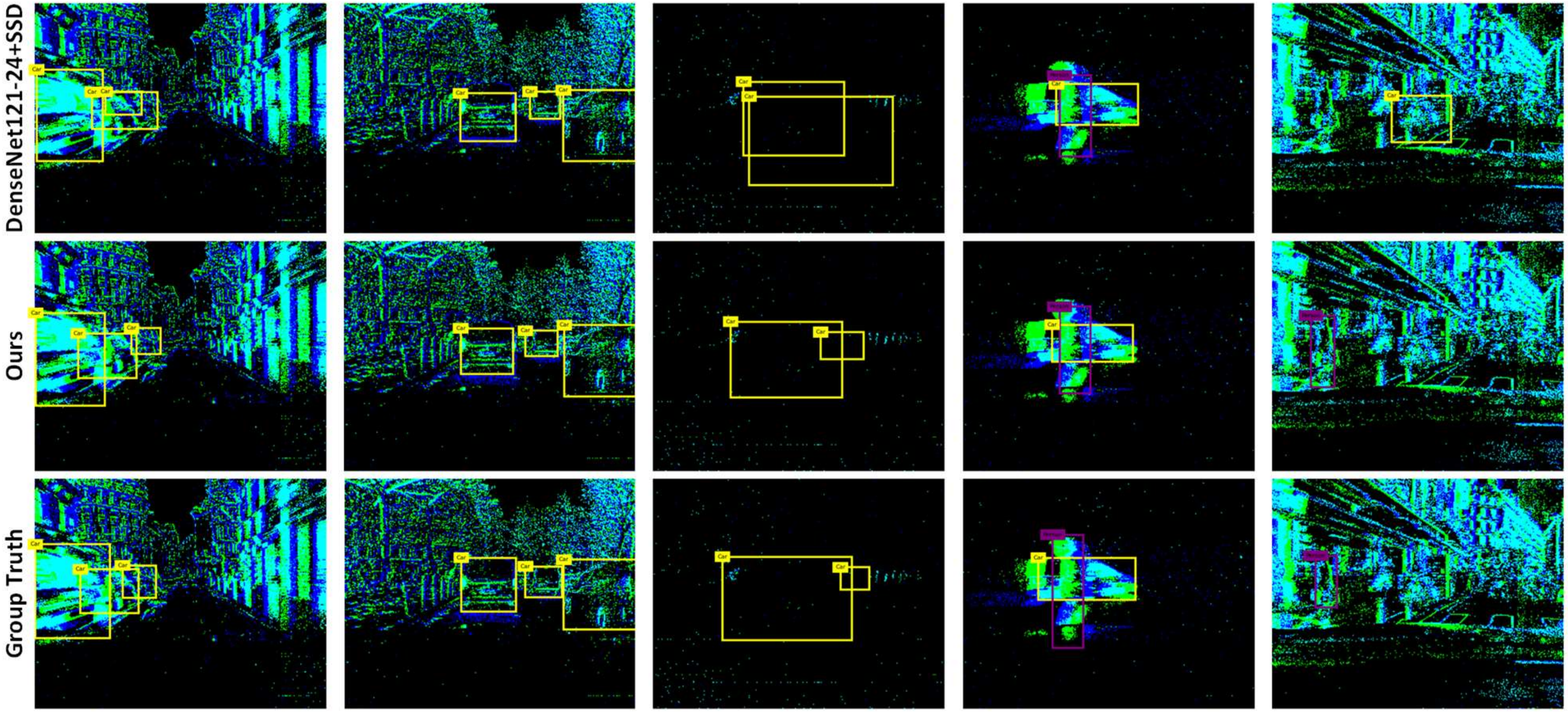

Fig6. Inference results of the model on the GEN1 dataset.

Conclusion

Summary and Future Outlook:

- **We propose Spiking Fusion Module, which is the first to implement spiking feature fusion in SNNs for event cameras.**
- **For the first time in SNNs applied to event cameras, we conduct a thorough study of different spiking decoding strategies and classification loss functions to determine their impact on model performance.**
- **On the GEN1 dataset, our SFOD achieves the state-of-the-art object detection performance of 32.1% mAP for SNN-based models.**
- **In the future, we believe that the performance of SFOD is expected to be further improved by adopting a more effective data augmentation strategy. It undeniably represents a promising research direction.**