Label-guided Real-time Fusion Network for RGB-T Semantic Segmentation

Zengrong Lin[1], Baihong Lin[2], Yulan Guo[1]

➤ https://github.com/llzeros/LRFNet-master

[1] School of Electronics and Communication Engineering, Sun Yat-sen University, Shenzhen, China
[2] Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China, Shenzhen, China

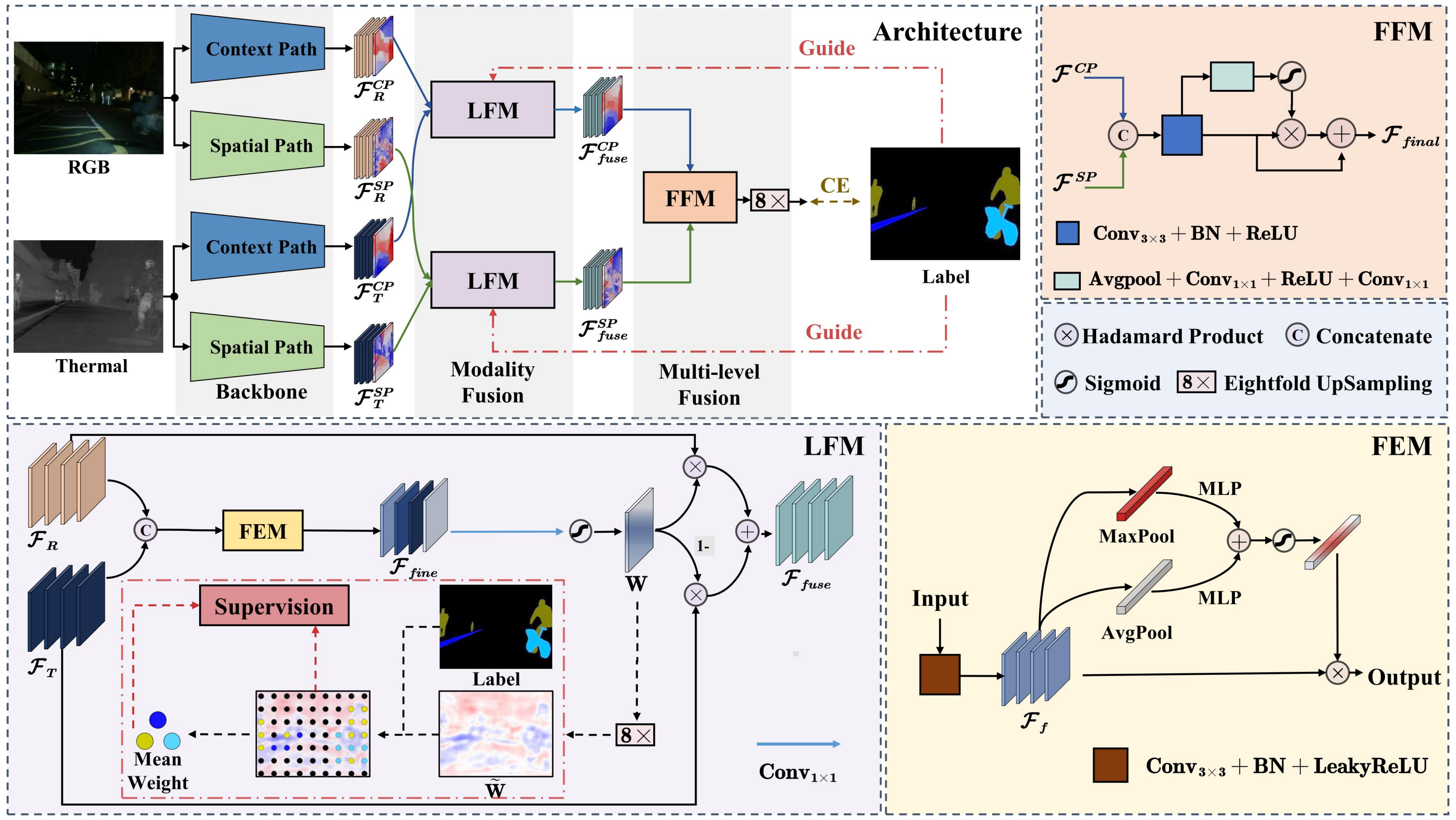
Background

- RGB-T semantic segmentation tries to accurately classify each pixel of a RGB image into a specific label by using a thermal image as complementary data.
- Most deep learning-based RGB-T methods suffer from low computational efficiency, i.e., they are not suitable for RGB-T real-time semantic segmentation.

Motivation

- Some lightweight networks with a good balance between accuracy and efficiency for segmentation in recent years having already adopted for RGB-based real-time semantic segmentation, but they are seldom discussed for RGB-T real-time semantic segmentation.
- Conventional fusion modules based on element-wise addition or concatenation fail to fully integrate information of paired RGB and thermal images.
- Most current fusion module designs are conducted based on heavy backbones, and their performance are not validated based on real-time lightweight backbones.

Method



Contribution

- This paper proposes a novel Label-guided Real-time Fusion Network which fuses detail and context features of RGB and thermal images extracted from double two-pathway lightweight backbones respectively based on the proposed Label-guided Fusion Module (LFM) to achieve fast and accurate perception.
- The LFM conducts weighted feature fusion based on a spatial attention map generated with the guidance of semantic label in the training phase to accurately indicate the contribution of different modalities.
- Our model achieves 55.1% mIoU with the speed of 111.3FPS on the MFNet dataset, and 78.4% mIoU with the speed of 67.3FPS on the PST900 dataset.

Reference

1. Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. Bisenet: Bilateral segmentation network for real-time semantic segmentation. ECCV, 2018.

2. Qiang Zhang, Shenlu Zhao, Yongjiang Luo, Dingwen Zhang, Nianchang Huang, and Jungong Han. Abmdrnet: Adaptive-weighted bi-directional modality difference reduction network for rgb-t semantic segmentation. CVPR, 2021.

3. Wujie Zhou, Shaohua Dong, Caie Xu, and Yaguan Qian. Edge-aware guidance fusion network for rgb–thermal scene parsing. AAAI, 2022.

Experimental Results

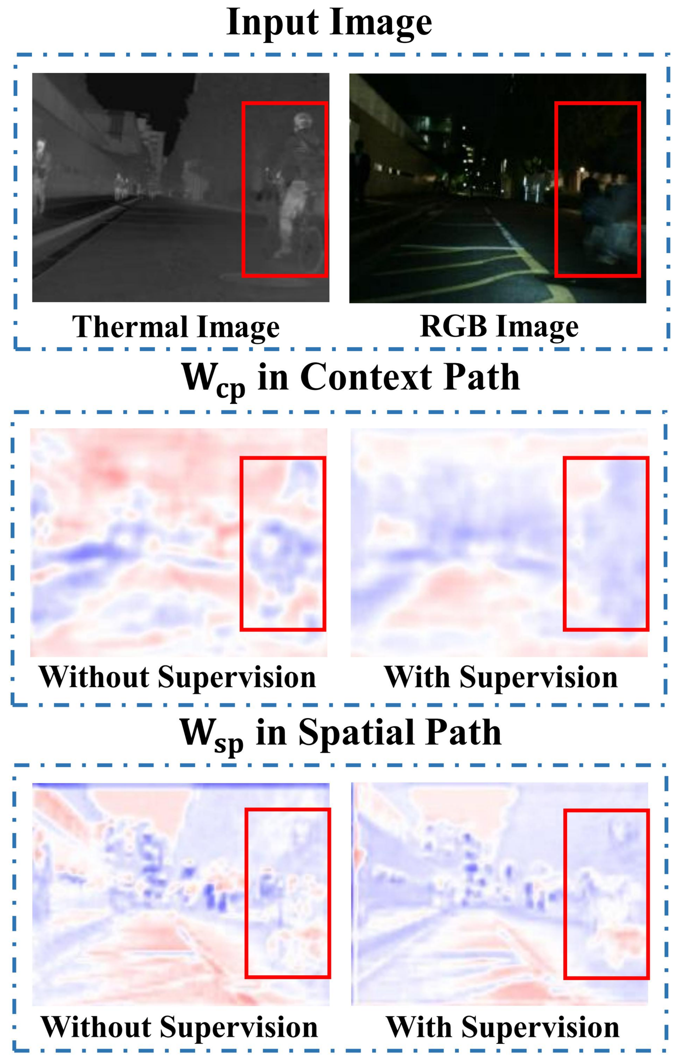
	Simple Fusion Methods (w/o LFM)			Placement of LFM	
Model	4 Channels	Concatenation	Addition	before FFM	after FFM
mAcc	64.7	65.7	65.5	68.0	64.5
mIoU	52.4	53.8	53.7	55.1	53.6
	Loss Function			Structure of FEM	
Model	CE	WCE	WCE+$\mathcal{L}(*,*)$	with FEM	w/o FEM
mAcc	61.2	66.7	68.0	68.0	67.1
mIoU	53.6	54.3	55.1	55.1	54.2

Ablation Study

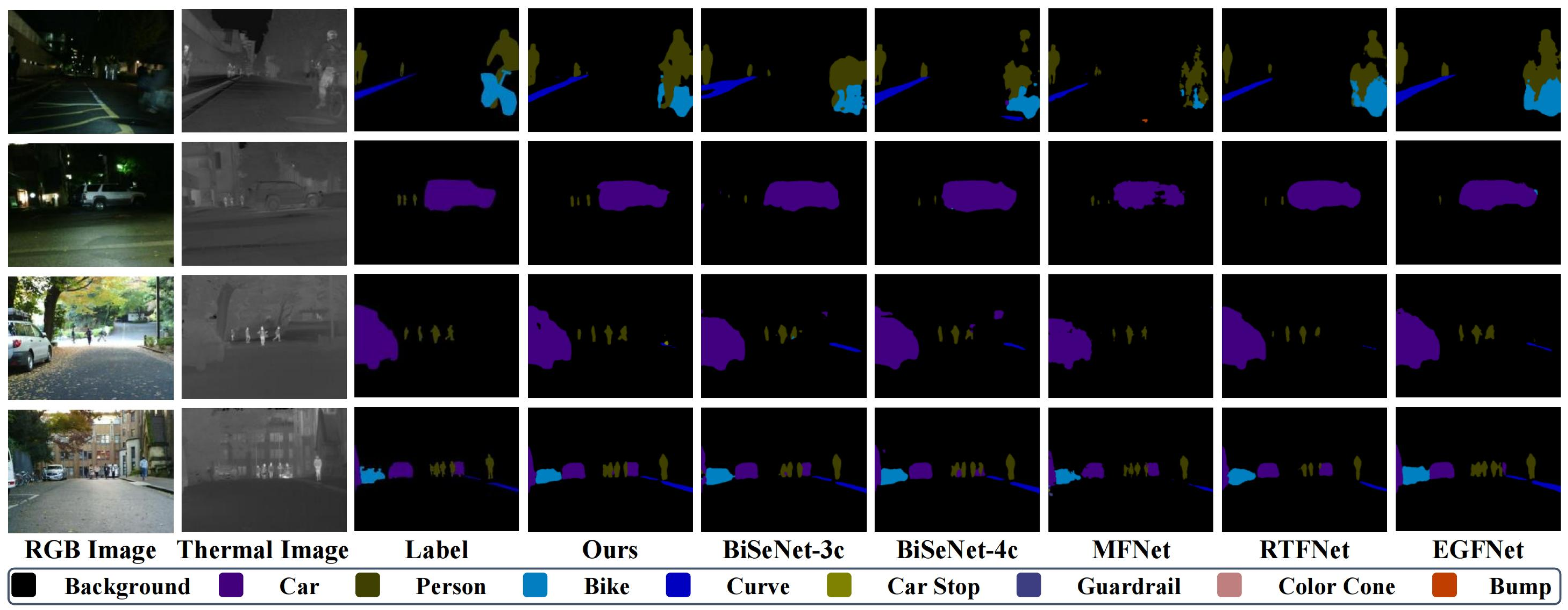
Methods	Type	Publication	Backbone	Params. (M)	FLOPs (G)	FPS	mAcc	mIoU
BiSeNet-3c	RGB	ECCV 2018	ResNet18	13.3	17.4	241.7	61.4	48.2
BiSeNet-4c	RGB	ECCV 2018	ResNet18	13.3	17.9	237.3	64.7	52.4
MFNet	RGBT	IROS 2017	No	0.7	8.4	178.1	45.1	39.7
RTFNet-152	RGBT	RAL 2019	ResNet152	254.5	290.3	16.4	63.1	53.2
FuseSeg	RGBT	T-ASE 2021	DenseNet161	100.1	141.0	20.5	70.6	54.5
ABMDRNet	RGBT	CVPR 2021	ResNet50	64.6	194.3	23.1	69.5	54.8
EGFNet	RGBT	AAAI 2022	ResNet101	62.8	201.3	20.5	**72.7**	54.8
Ours	RGBT	-	ResNet18	**25.9**	**32.0**	**111.3**	68.0	**55.1**

Comparison results from the MFNet dataset

Visualization Examples



Spatial Attention Maps



Visualization segmentation examples of our method and five representative state-of-the-art methods