1.9 Localized Semantic Feature Mixers for Efficient Pedestrian Detection in Autonomous Driving

Abdul Hannan Khan | DFKI

Detection

Abstract

Pedestrian Detection is vital for autonomous driving. It needs to be performant and efficient to provide effective and in-time perception. The existing approaches compromise the efficiency to improve performance by deploying large, complex and computationally expensive deep learning approaches. We revisit the architecture of pedestrian detection networks, specially feature pyramid network and detection head, to remove the components which are computationally expensive and do not contribute directly to the learning capability of the network. We design novel detector which uses MLP-Mixers to boost performance. We benchmark our proposed method on 4 well-established pedestrian detection datasets and report our findings.

Objectives

- Improve efficiency and performance of pedestrian detectors, specially in small and heavily occluded cases.
- Reformulate the feature pyramid pooling design to remove expensive operations.

Methodology

- Super Pixel Pyramid Pooling (SP3)
  - Combines patches from different backbone stages into unified representation called superpixels.
  - Single fully-connected layer for feature enrichment and filtering.
- Dense Focal Detection Network (DFDN)
  - Anchor-free design
  - MLP-Mixers [1] blocks to boost performance.
  - Works on patches to improve efficiency and better local information flow.
- Hard-Mixup augmentation to boost performance in small and occluded cases.

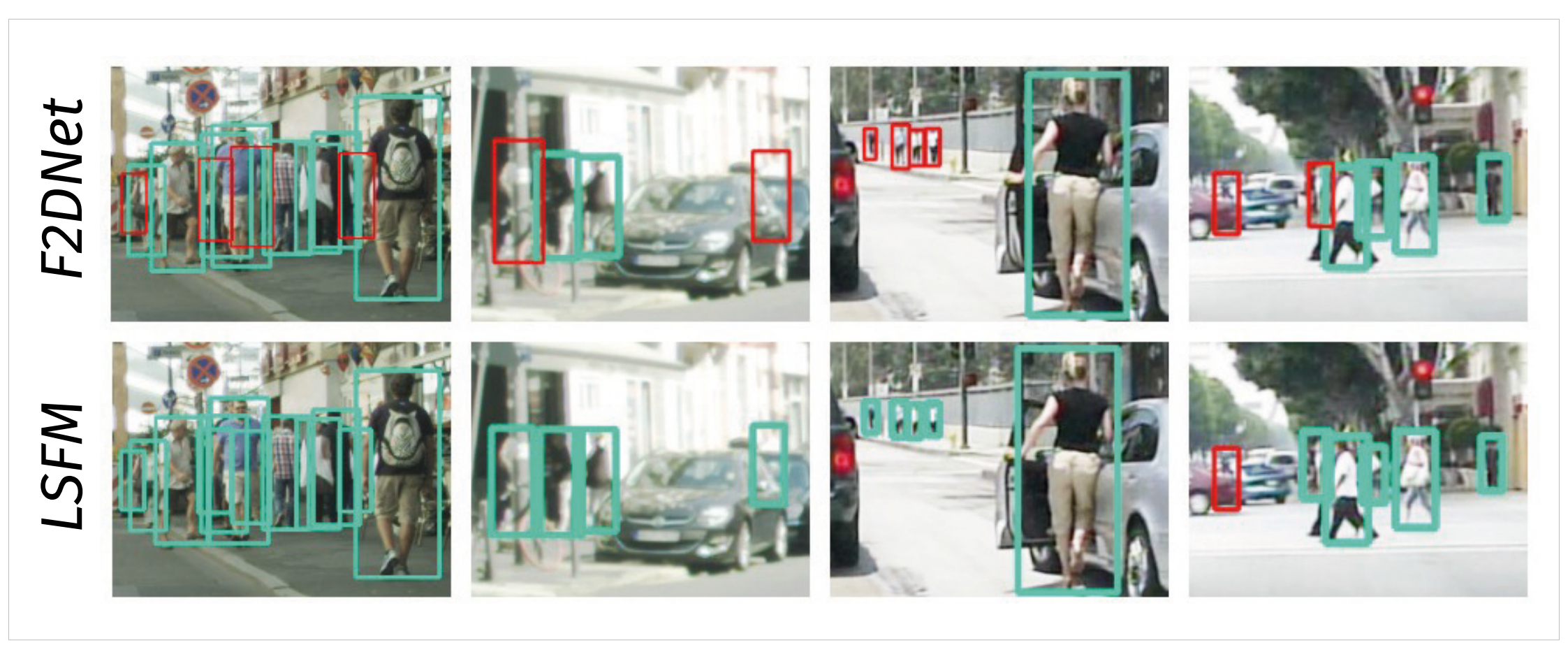

Figure 1: Qualitative Comparison. (© DFKI GmbH)

Metrics

- Log-Average Miss Rate (LAMR):

$$\exp\left(\frac{1}{9}\Sigma_f \log\left(mr\left(\underset{fppi(c)<f}{\operatorname{argmax}} fppi(c)\right)\right)\right)$$

Results

- Caltech

Method	Reasonable	Small	Heavy	Inference
Pedestron [2]	2.6	2.8	24.4	0.20s
F2DNet [3]	1.2	1.4	19.6	0.14s
LSFM (ours)	**1.0**	**0.2**	**19.5**	**0.09s**

- City Persons

Method	Reasonable	Small	Heavy	Inference
Pedestron [2]	8.9	10.6	29.6	0.73s
F2DNet [3]	6.8	9.0	26.0	0.44s
LSFM (ours)	**6.7**	**6.7**	**23.5**	**0.18s**

- Euro City Persons

Method	Reasonable	Small	Heavy	Inference
F2DNet [3]	6.0	11.1	29.1	0.41s
Pedestron [2]	4.7	10.2	24.7	0.44s
LSFM (ours)	**4.1**	**9.5**	**20.9**	**0.17s**

- Beats human baseline on Caltech dataset
- Beats SOTA with **55 %** lesser inference time

References

[1] Tolstikhin, Ilya O., et al. Mlp-mixer: An all-mlp architecture for vision. *NIPS*, 2021.

[2] Hasan, Irtiza, et al. Generalizable pedestrian detection: The elephant in the room. *CVPR* 2021.

[3] Khan, Abdul Hannan, et al. F2DNet: Fast focal detection network for pedestrian detection. *ICPR* 2022.

City Persons
https://www.v7labs.com/open-datasets/citypersons

Caltech Pedestrians
https://data.caltech.edu/records/f6rph-90m20

Euro City Persons
https://eurocity-dataset.tudelft.nl/

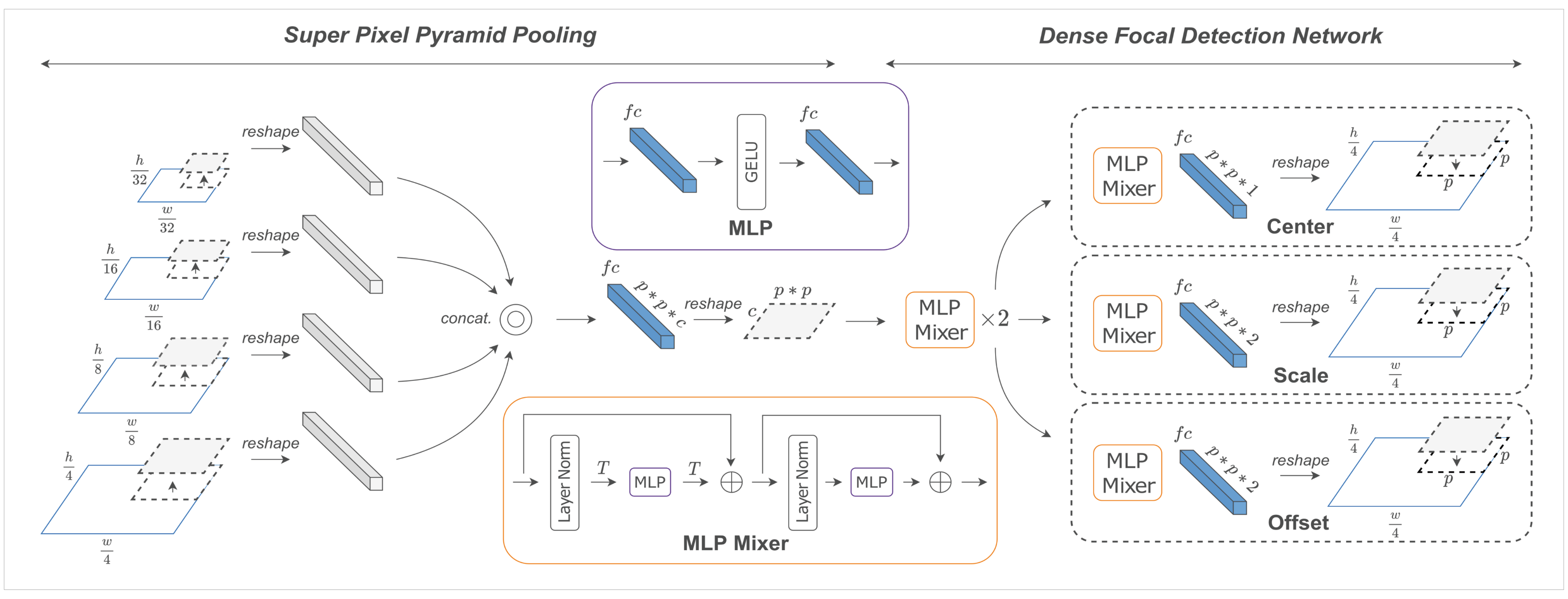

Figure 2: Model architecture of Localized Semantic Feature Mixers. (© DFKI GmbH)

Partners

External partners

For more information contact:
hannan.khan@dfki.de

KI Wissen is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

www.kiwissen.de   @KI_Familie   KI Familie

Funded by the European Union
NextGenerationEU

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag