KI WISSEN

Automotive AI Powered by Knowledge

KI Wissen Final Event | 21-22 March 2024

Efficient Pedestrian Detection with Inter-stage Knowledge Integration

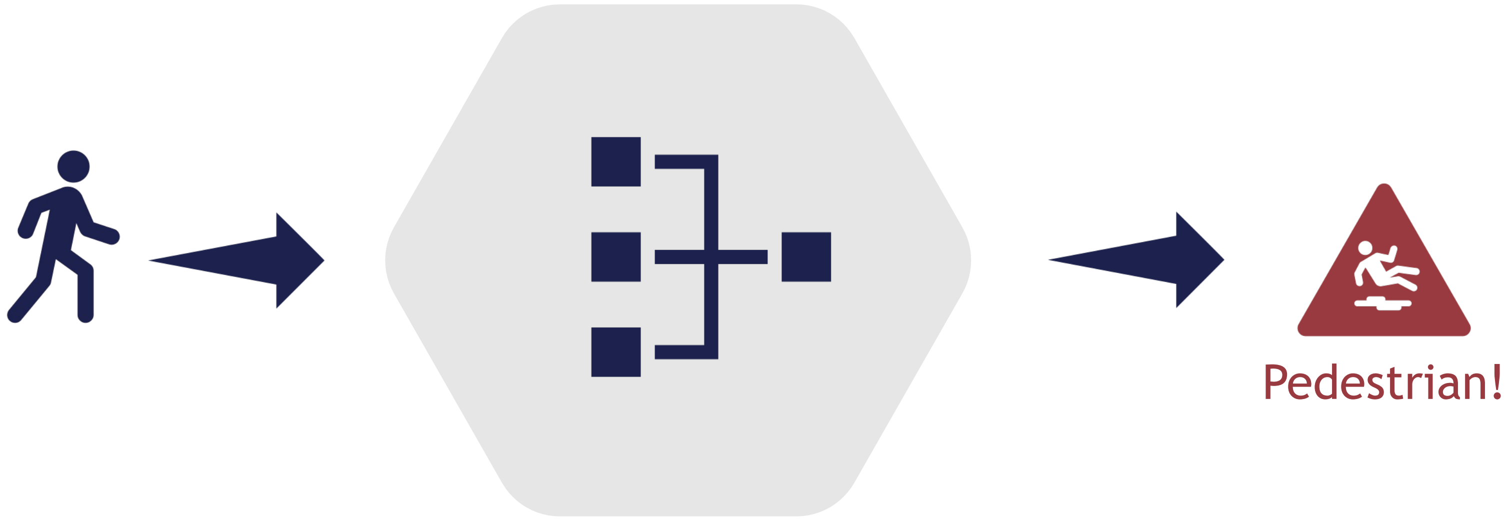
Abdul Hannan Khan | DFKI

1

Pedestrian Detection

Pedestrian Detection: The Problem

- Detect and Localize pedestrians in a given scene
- Not only limited to autonomous driving
- Tolerate riders, seated pedestrians and reflections



Pedestrian Detection: Challenges

- Challenges
  - Heavy Occlusions
  - Motion Blur
  - Higher Inference Time

Crowd

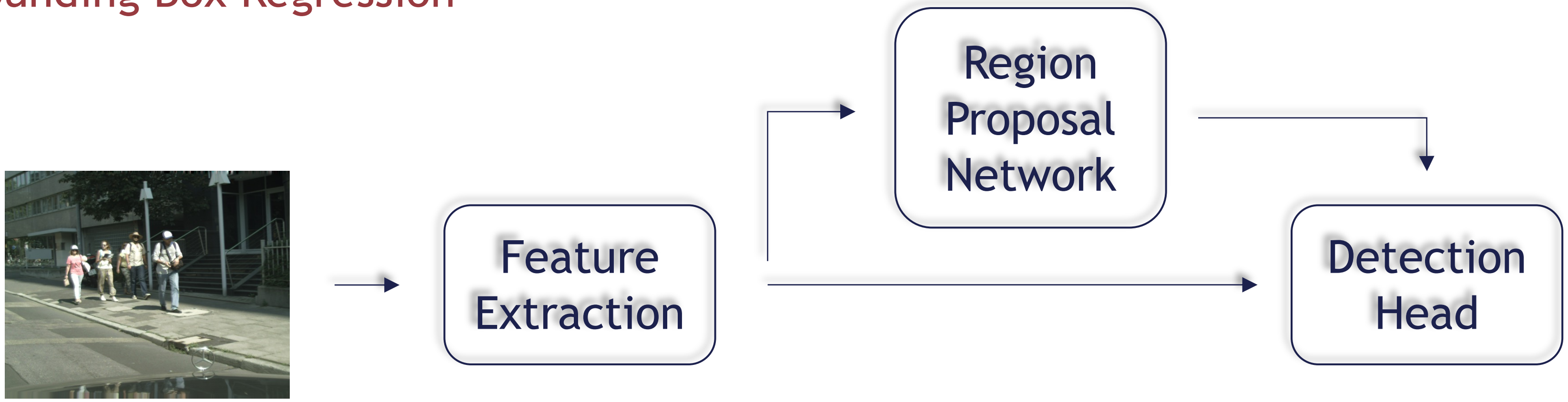
Heavy Occlusion

Pedestrian Detection: Datasets

Dataset	Images	Pedestrians	Resolution
Caltech Pedestrian[4]	42,782	13,674	640 x 480
City Persons[5]	2,975	19,238	2048 x 1024
Euro City Persons[6]	21,795	201,323	1920 x 1024

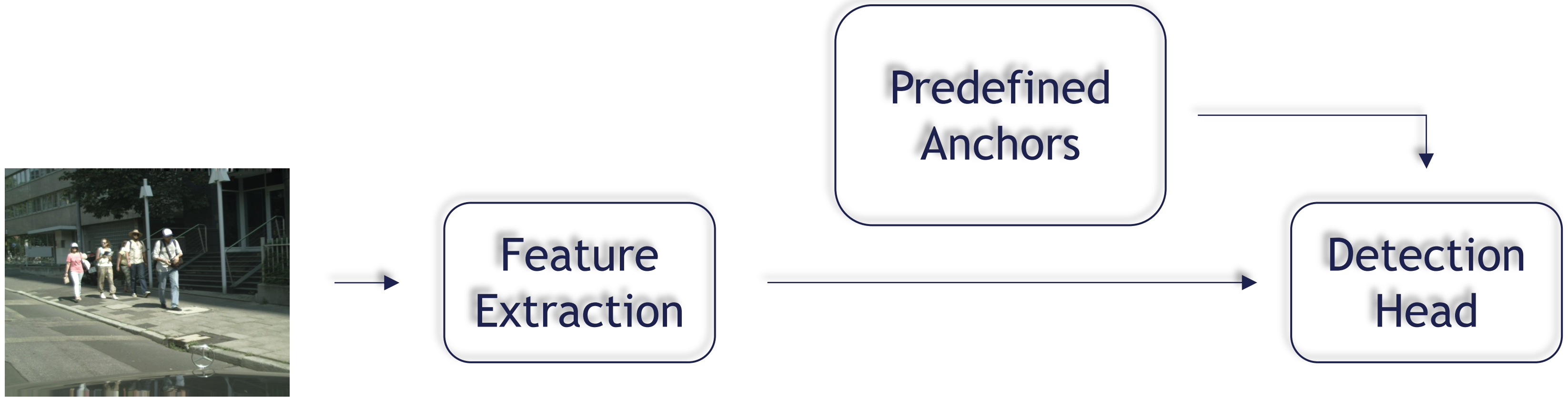
Existing Solutions: Two Stage Architectures

- Higly Performant
- Computationally Expensive
- Redundant Bounding Box Regression



Existing Solutions: Single Stage Architectures

- Faster than two-stage architectures
- Performance drop



Existing Solutions: Anchor Free Architectures

- No Achors
- Hard Centers are hard to learn

Feature Extraction → Detection Head

Existing Solutions: Archor Free Architectures

Gaussian based Soft Centres

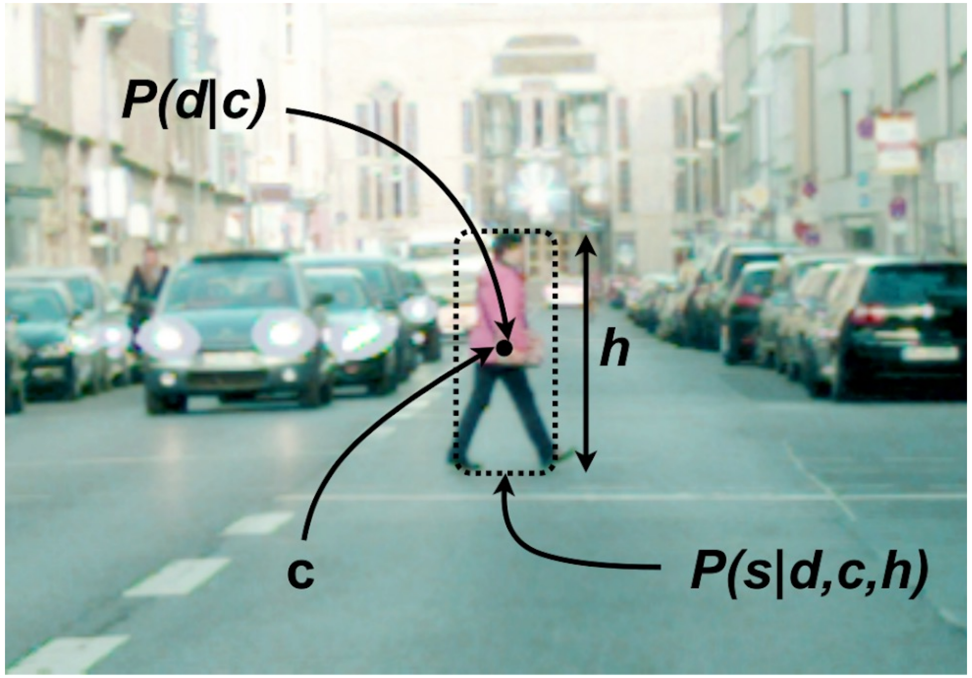
Increased False Positives!

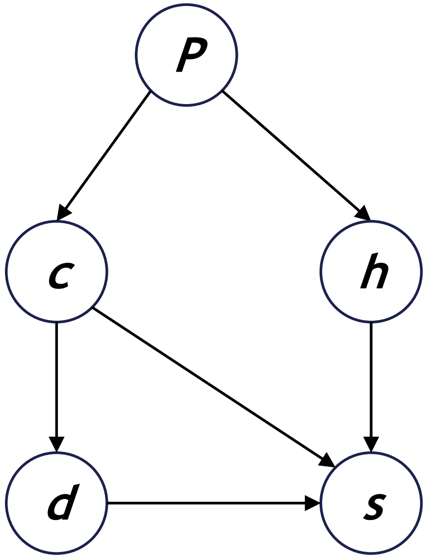
*Image is taken from CSP[2]

2

F2DNet: Fast Focal Detection Network

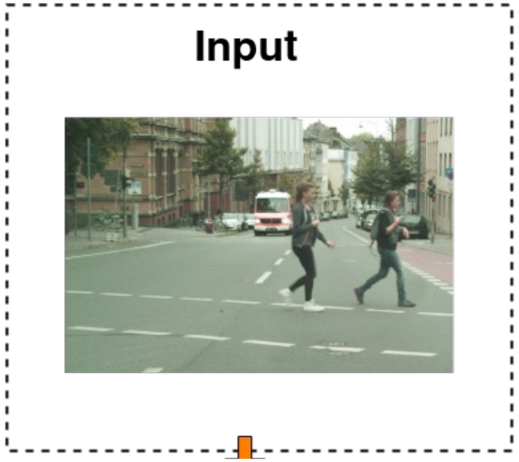
F2DNet: Two Stage & Anchor Free





$$P(\neg s, d|c, h) = P(\neg s|d, c, h)P(d|c)$$

F2DNet: Two Stage & Anchor Free



F2DNet: Efficiency and Performance

- Better results compared to Pedestron
- The time is reported on Nvidia GTX-1080Ti
- F2DNet takes on average ~28% lesser time compared to Cascade R-CNN[1]

City Persons

Method	Reasonable	Small	Heavy	Inference
Pedestron[1]	11.2	14.0	37.0	0.73s
BGCNet[3]	8.8	11.6	43.9	-
F2DNet	**8.7**	**11.3**	**32.6**	**0.44s**

Caltech

Method	Reasonable	Small	Heavy	Inference
Pedestron[1]	6.2	7.4	55.3	0.20s
CSP[2]	5.0	6.8	46.6	-
F2DNet	**2.2**	**2.5**	**38.7**	**0.14s**

ECP

Method	Reasonable	Small	Heavy	Inference
Pedestron[1]	6.6	13.6	33.3	0.44s
F2DNet	**6.1**	**10.7**	**28.2**	**0.41s**

Measure: MR^{-2}
Lower is better

3

LSFM: Localized Semantic Feature Mixers

LSFM: Localized Semantic Feature Mixers

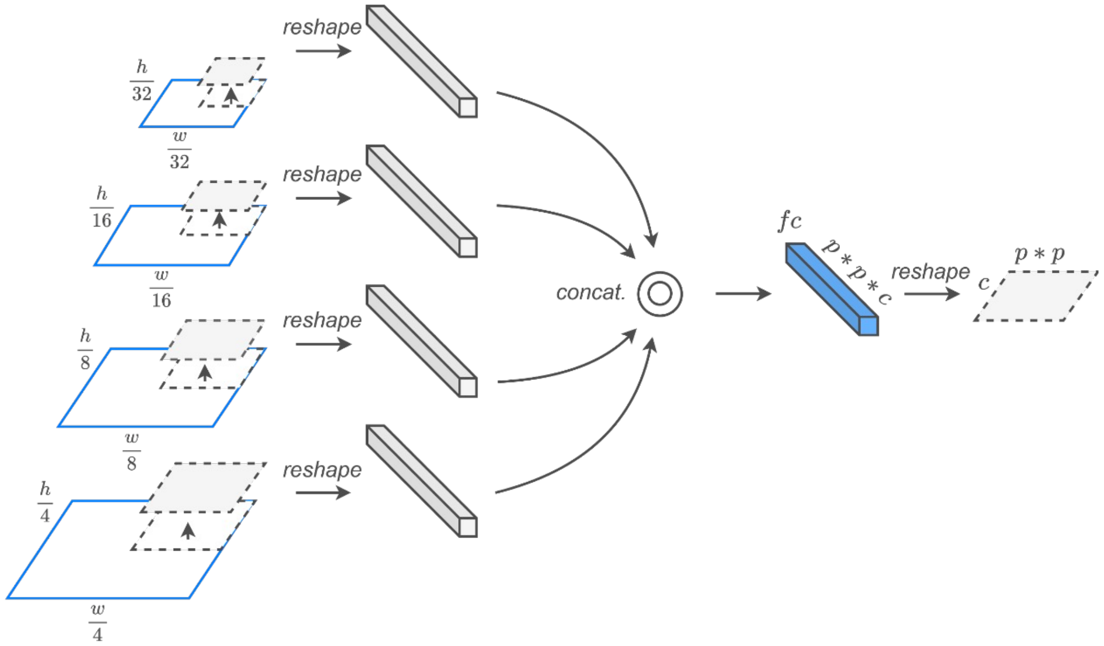
- MLPMixers[7] based pedestrian detection architecture
- Uses highly efficient feature enrichment neck
- Uses high-level semantic feature representation of pedestrians
- Works with batches of patches (super pixels) to improve local information flow and increase cache efficiency

1.9

Poster

LSFM: Super Pixel Pyramid Pooling

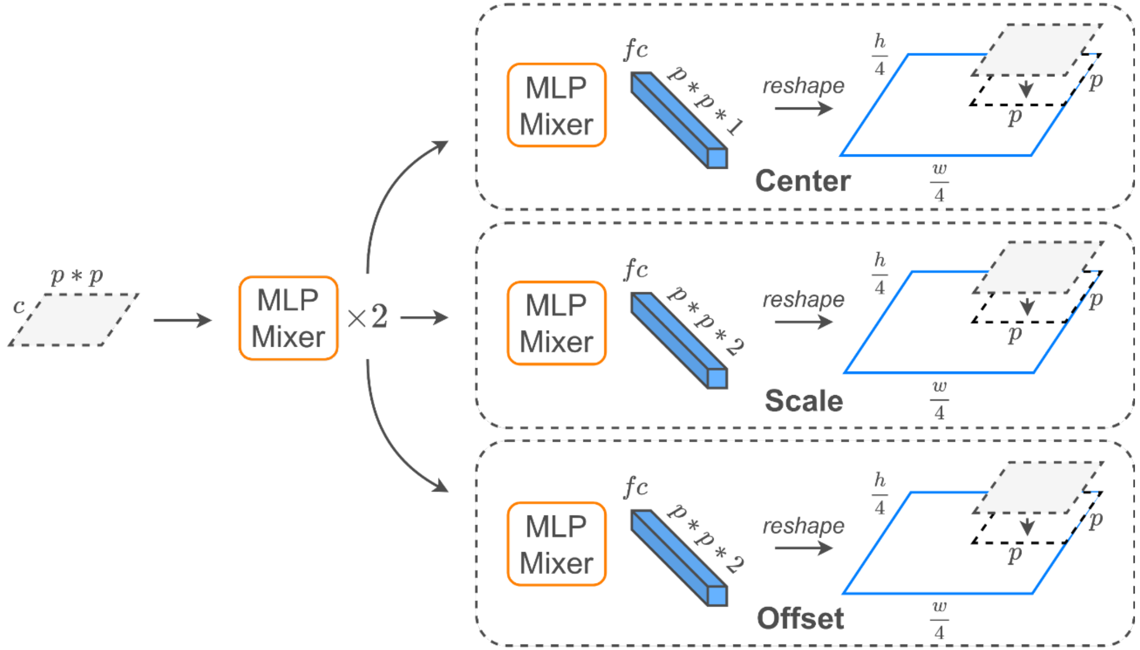
- Combines patches from different backbone stages into unified representation called Super Pixels
- Single fully-connected layer for feature enrichment and filtering
- Performant and cache efficient



1.9
Poster

LSFM: Dense Focal Detection Network

- Anchor-free detection head
- MLPMixers[7] blocks to boost performance
- Works on patches to improve efficiency and boost local information flow



1.9
Poster

LSFM: ConvMLP Pin Backbone & Hard Mixup Augmentation

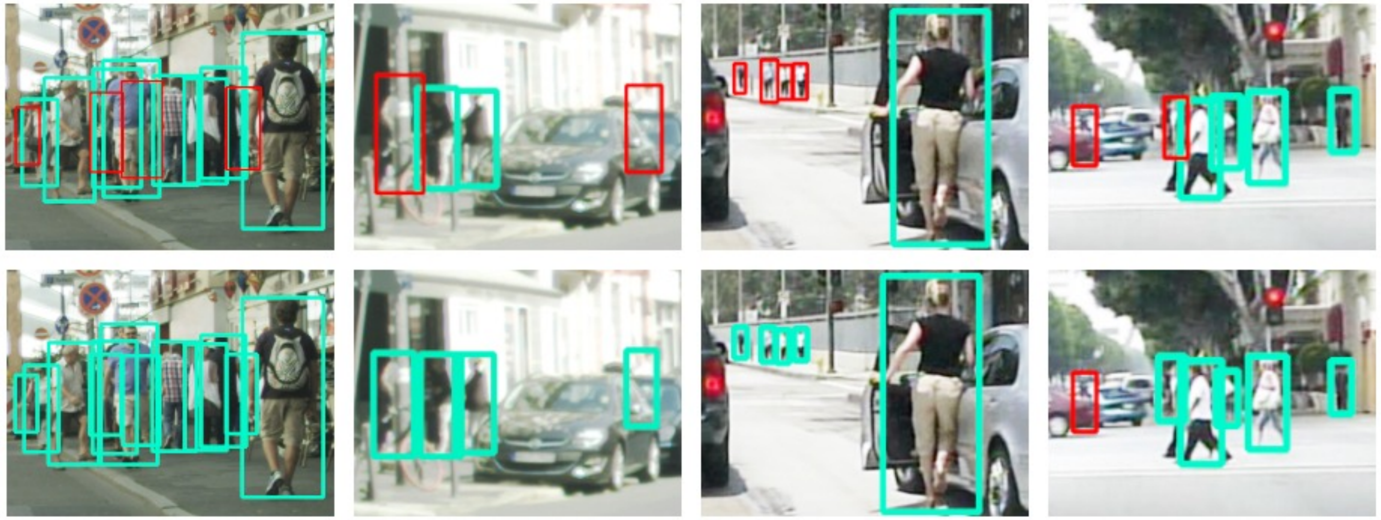
- ConvMLP Pin
  - Based on ConvMLP[8]
  - Uses MLPMixers[7] with convolutions to be applicable for variable sized input
  - Deep yet not wide backbone to learn high-level semantic features while being efficient
- Hard Mixup Augmentation
  - To boost performance in small and heavily occluded cases

1.9

Poster

LSFM: Qualitative Comparison

- Cyan shows true positives & red indicate false negatives
- LSFM performs significantly better than F2DNet
- Some hard cases where LSFM misses pedestrian as well
- Few very rare cases where F2DNet detects pedestrian while LSFM misses



1.9
Poster

LSFM: Quantitative Comparison

- Beats SOTA in popular pedestrian datasets
- Beats human baseline on Caltech dataset[4]
- 55% lesser inference time

City Persons

Method	Reasonable	Small	Heavy	Inference
Pedestron[1]	8.9	10.6	29.6	0.73s
F2DNet	6.8	9.0	26.0	0.44s
LSFM	**6.7**	**6.7**	**23.5**	**0.18s**

Caltech

Method	Reasonable	Small	Heavy	Inference
Pedestron[1]	2.6	2.8	24.4	0.20s
F2DNet	1.2	1.4	19.6	0.14s
LSFM	**1.0**	**0.2**	**19.5**	**0.09s**

ECP

Method	Reasonable	Small	Heavy	Inference
F2DNet	6.0	11.1	29.1	0.41s
Pedestron[1]	4.7	10.2	24.7	0.44s
LSFM	**4.1**	**9.5**	**20.9**	**0.17s**

Measure: MR^{-2}
Lower is better

4

LSFM for Traffic Object Detection

Traffic Object Detection

- Traffic actors belong to multiple classes, although pedestrians are most risky, collision with other objects must be avoided as well.
- Due to increased number of constraints, architectures which perform well for pedestrian detection should generalize well to other objects.
- Existing object detectors are performant but far away from being real-time which is critical for autonomous driving.

Traffic Object Detection

- Extend state-of-the-art pedestrian detection model LSFM to enable multiclass object detection.
- Instead of predicting pedestrian or background predict K class probabilities.
- Class normalized focal loss instead of class agnostic instance normalized focal loss.

$$L_{center} = \frac{1}{C} \sum_{c} \frac{1}{K_c} \sum_{t} \alpha_c(t) FL_c(p_t, y_t)$$

Traffic Object Detection Results

- Beats state-of-the-art object detectors with significant margin
- Inference time is based on RTX 3090 with single sample per batch

TJU-Traffic [9]

Method	mAP	mAP50	mAP75	fps	RTOP
Cascade RCNN	57.9	82.7	66.6	6.7	33.8
LSFM	**60.4**	85.7	70.0	11.2	39.1
YOLOv3	56.8	85.4	64.1	14.9	40.1
LSFM P	56.9	83.7	64.4	**33.3**	**56.9**

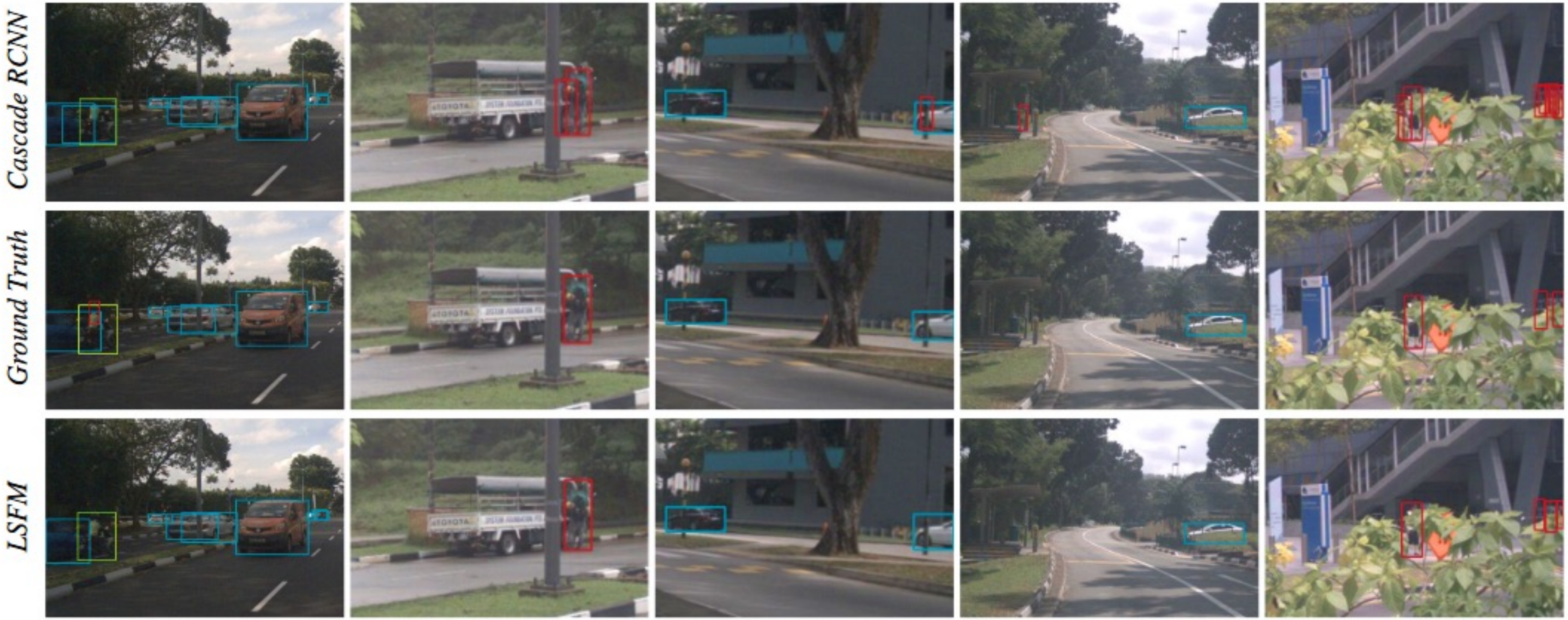
NuImage [10]

Method	mAP	mAP50	mAP75	fps	RTOP
Cascade RCNN	47.9	-	-	12.1	31.7
LSFM	**48.1**	76.2	51.9	14.3	33.5
YOLOv3	41.8	71.1	43.0	20.5	33.6
LSFM P	46.1	74.6	48.7	**30.3**	**46.1**

BDD100K [11]

Method	mAP	mAP50	mAP75	fps	RTOP
Cascade RCNN	**32.4**	-	-	14.3	22.6
LSFM	31.5	59.1	29.0	17.4	23.6
YOLOv3	27.5	54.5	23.8	32.4	27.5
LSFM P	28.2	55.7	24.4	**32.6**	**28.2**

Traffic Object Detection Results



5

Demonstration

Demonstration

- AVL AD Stack
  - Carla
  - Object Detection
    1. LSFM -> 2D detections
    2. Capgemini solution -> 3D detections
    3. Published to -> AD Stack

6

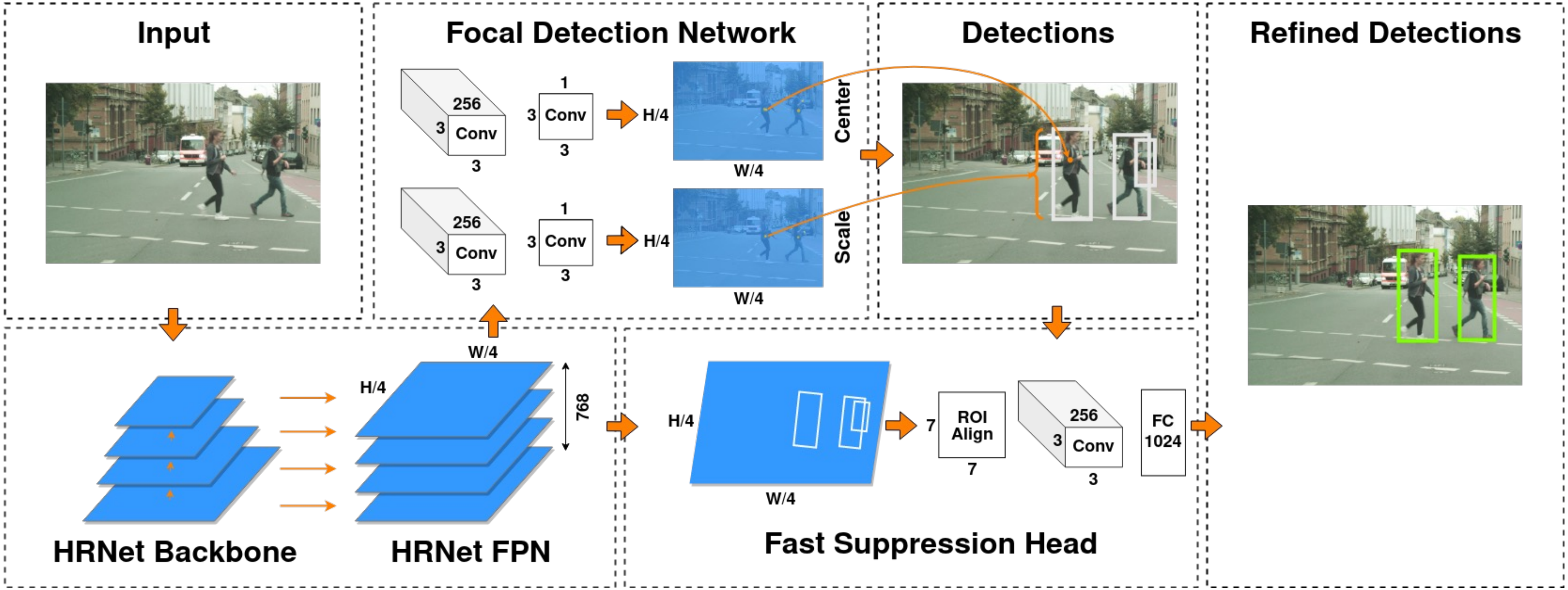
Knowledge Building Blocks

Knowledge Building Blocks

ID	Knowledge Description (Machine Readable)	Knowledge Representation (Human Readable)	Use Case (Operational Design Domain (ODD))	Integration Method
~KB0039	Object & Environment Interaction resulting in Silhouette & Gradients	Object Contours	Traffic Object Detection	Networks Inside Network

Knowledge Building Blocks

- Inter-stage Knowledge



KI WISSEN

Automotive AI Powered by Knowledge

Thank You!

References

F2DNet | Khan, Abdul Hannan et al. F2DNet: Fast Focal Detection Network for Efficient Pedestrian Detection. ICPR 2022.

LSFM | Khan, Abdul Hannan et al. Localized Semantic Feature Mixers for Efficient Pedestrian Detection in Autonomous Driving. CVPR 2023.

[1]Hasan, Irtiza et al. Generalizable Pedestrian Detection: The Elephant In The Room. CVPR 2021.

[2]Liu, Wei et al. High-Level Semantic Feature Detection: A New Perspective for Pedestrian Detection. CVPR 2019.

[3]Li, Jinpeng et al. Box Guided Convolution for Pedestrian Detection. ACM MM 2020.

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

[8]Li, Jiachen, et al. Convmlp: Hierarchical convolutional mlps for vision. arXiv 2021.

[4]Caltech Pedestrians

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

[5]City Persons

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

[6]Euro City Persons

https://eurocity-dataset.tudelft.nl/

[9]TJU DHD Traffic

https://github.com/tjubiit/TJU-DHD

[10]NuImages

https://www.nuscenes.org/nuimages

[11]BDD100K

https://www.bdd100k.com/

KI WISSEN

Automotive AI Powered by Knowledge

Abdul Hannan Khan| DFKI | 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.

KI FAMILIE

VDA LEITINITIATIVE

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

www.kiwissen.de @KI_Familie KI Familie