KI WISSEN
Automotive AI Powered by Knowledge

KI Wissen Final Event | 21-22 March 2024

Knowledge-augmented Pedestrian Detection

Julian Wörmann | fortiss

KI WISSEN

Automotive AI Powered by Knowledge

Attention is all you need!

1.1 Poster

1.6 Poster

Motivation, Concepts and Goals

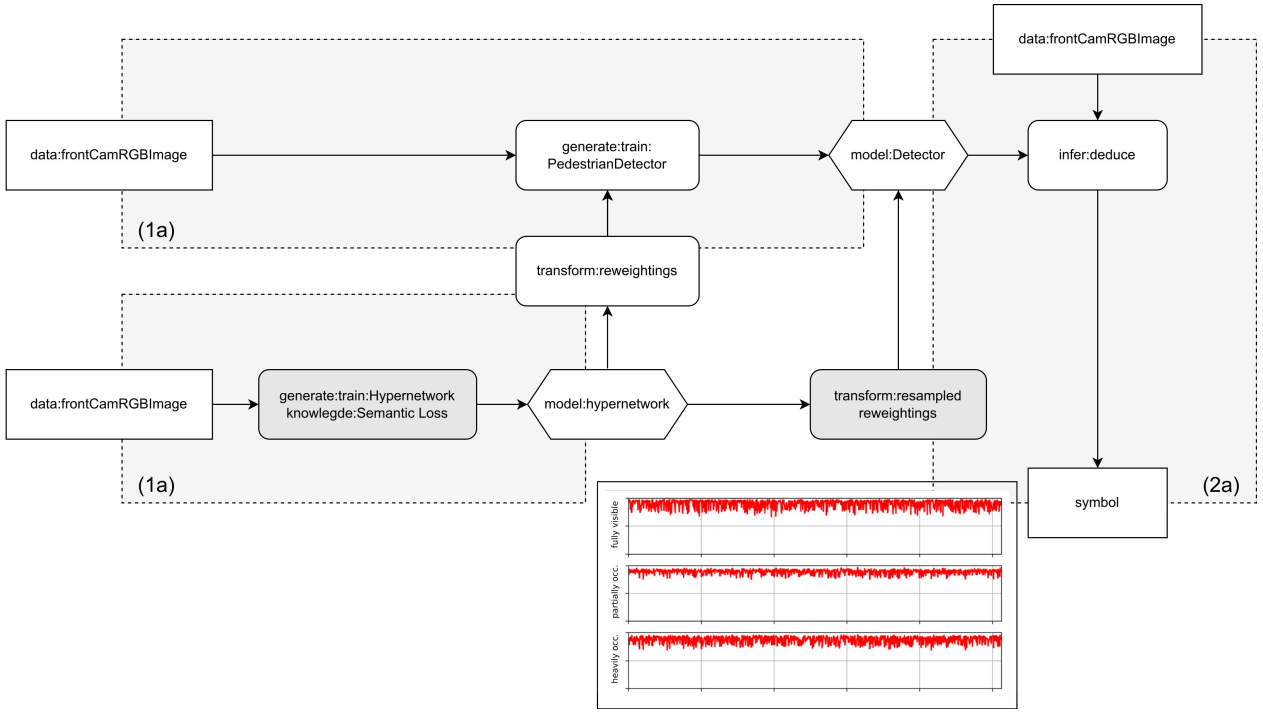
- Guide existing models towards reliable features for visual pedestrian detection
- Disentanglement of features; Association of semantics
- Utilization of concept activations extracted from support images
- Robust and explainable detection of occluded pedestrians from limited data

Knowledge Guided YOLO Detector

Training stage

- Query image is complemented with support images from sub-classes.
- Parallel hypernetwork determines reweighting coefficients

‘fully visible’
‘partially occluded’
‘heavily occluded’



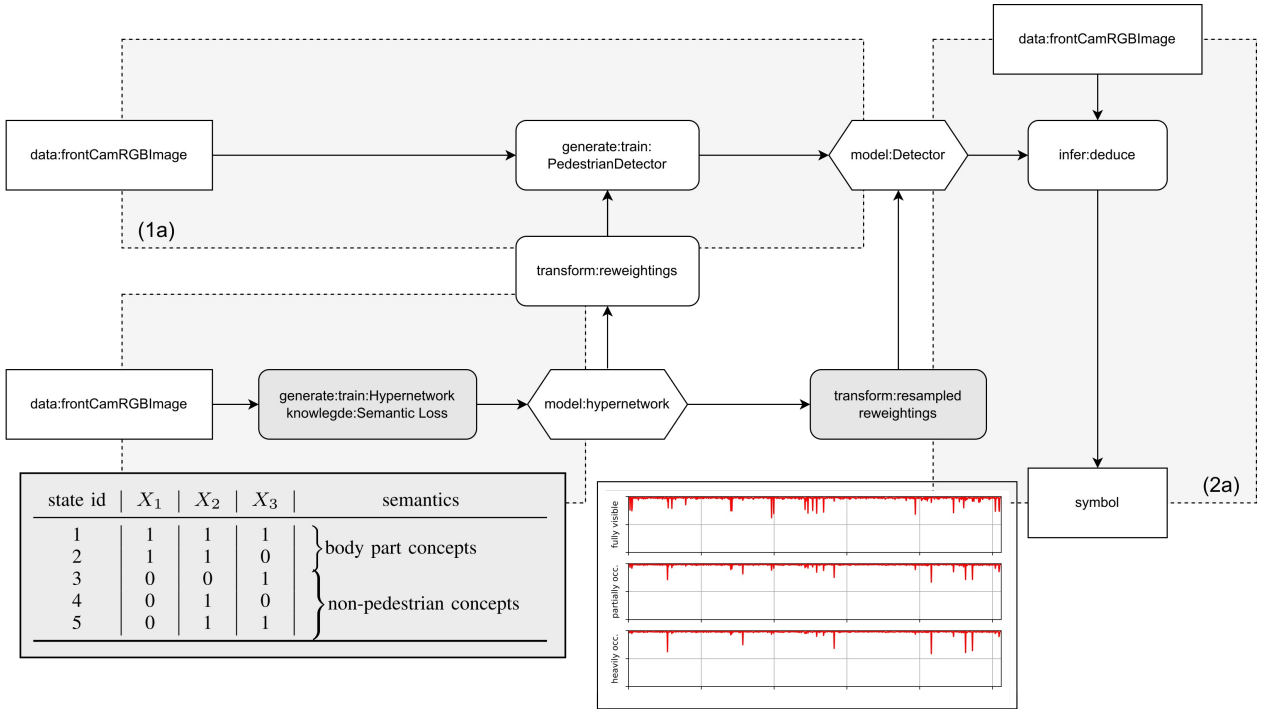
Knowledge Guided YOLO Detector

Training stage

- Query image is complemented with support images from sub-classes.
- Parallel hypernetwork determines reweighting coefficients
- Semantic loss enforces activations across sub-classes

$$\min_{\theta,\psi,\eta} \sum_i \left[L_d(I_i, \mathcal{S}_i) + \lambda \sum_l L_s(\alpha, \Omega_i^l) \right]$$

‘fully visible’
‘partially occluded’
‘heavily occluded’



Knowledge Guided YOLO Detector

Inference stage

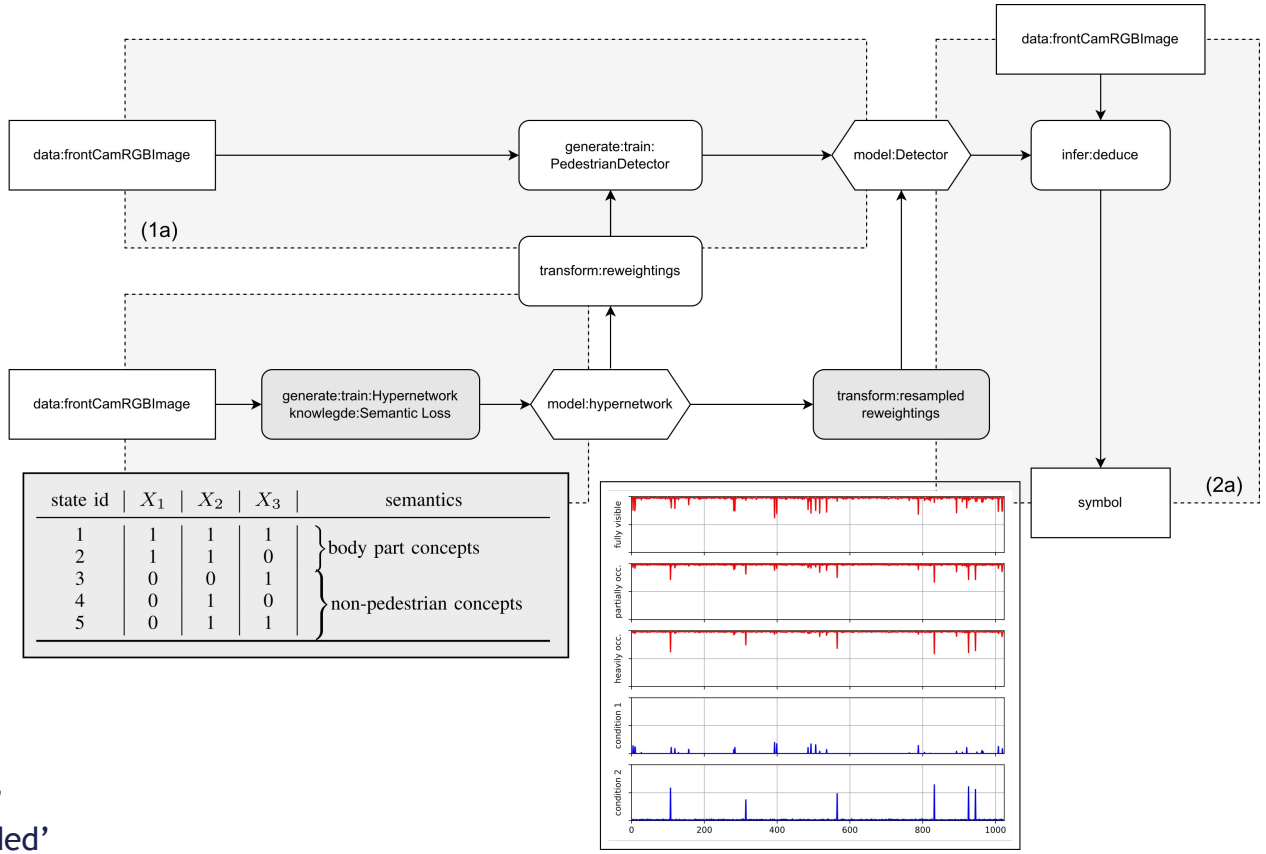
- Attenuation of feature maps based on conformance to semantics

$$c_1^l = \max\left(0, (1 - \bar{\omega}_1^l) + \min\left(\bar{\omega}_2^l + \bar{\omega}_3^l, 1\right) - 1\right)$$
$$c_2^l = \max\left(0, \bar{\omega}_1^l + \min\left((1 - \bar{\omega}_2^l) + (1 - \bar{\omega}_3^l), 1\right) - 1\right)$$

- Determine attenuated reweightings via

$$\hat{\omega} = \mathbb{1} + \gamma_1 \mathbf{c}_1 + \gamma_2 \mathbf{c}_2.$$

‘fully visible’
‘partially occluded’
‘heavily occluded’



Knowledge Guided YOLO Detector

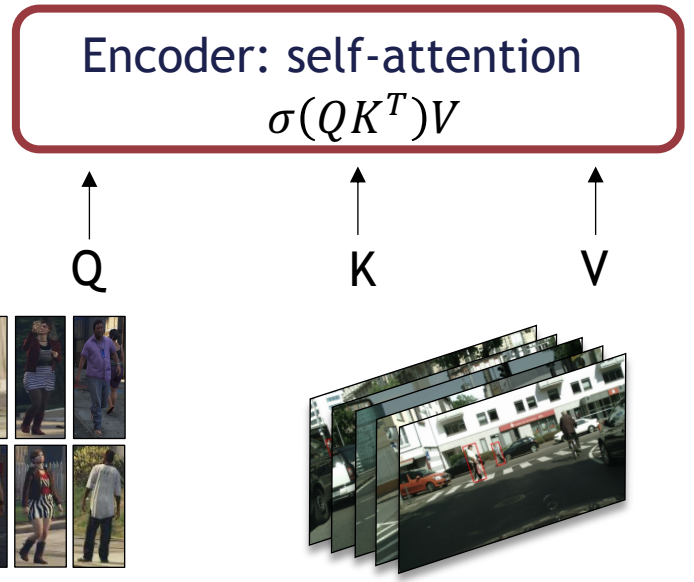
- During inference, model can be used with single image input, i.e., no computational overhead
- Depending on occlusion level, attenuation factors can be adjusted

- Potential for ensemble methods for addressing different levels of occlusion

Dataset	Split	Baseline	Knowledge Guided		
γ_2			0.0	−1.0	−2.0
Citypersons	vis + occ	**0.4350**	0.4065	0.3874	0.2723
	part. occ	0.3428	0.3035	**0.3445**	0.2953
	heavy occ	0.1362	0.1219	0.1935	**0.2761**
JAAD	vis + occ	**0.5081**	0.4963	0.3945	0.1853
	part. occ	0.2129	0.1968	**0.2598**	0.1964
	heavy occ	0.0897	0.0529	**0.1303**	0.1297
SHIFT	vis + occ	**0.3153**	0.2905	0.1815	0.1122
	part. occ	**0.2057**	0.1513	0.1712	0.1196
	heavy occ	**0.1322**	0.0595	0.0987	0.1006

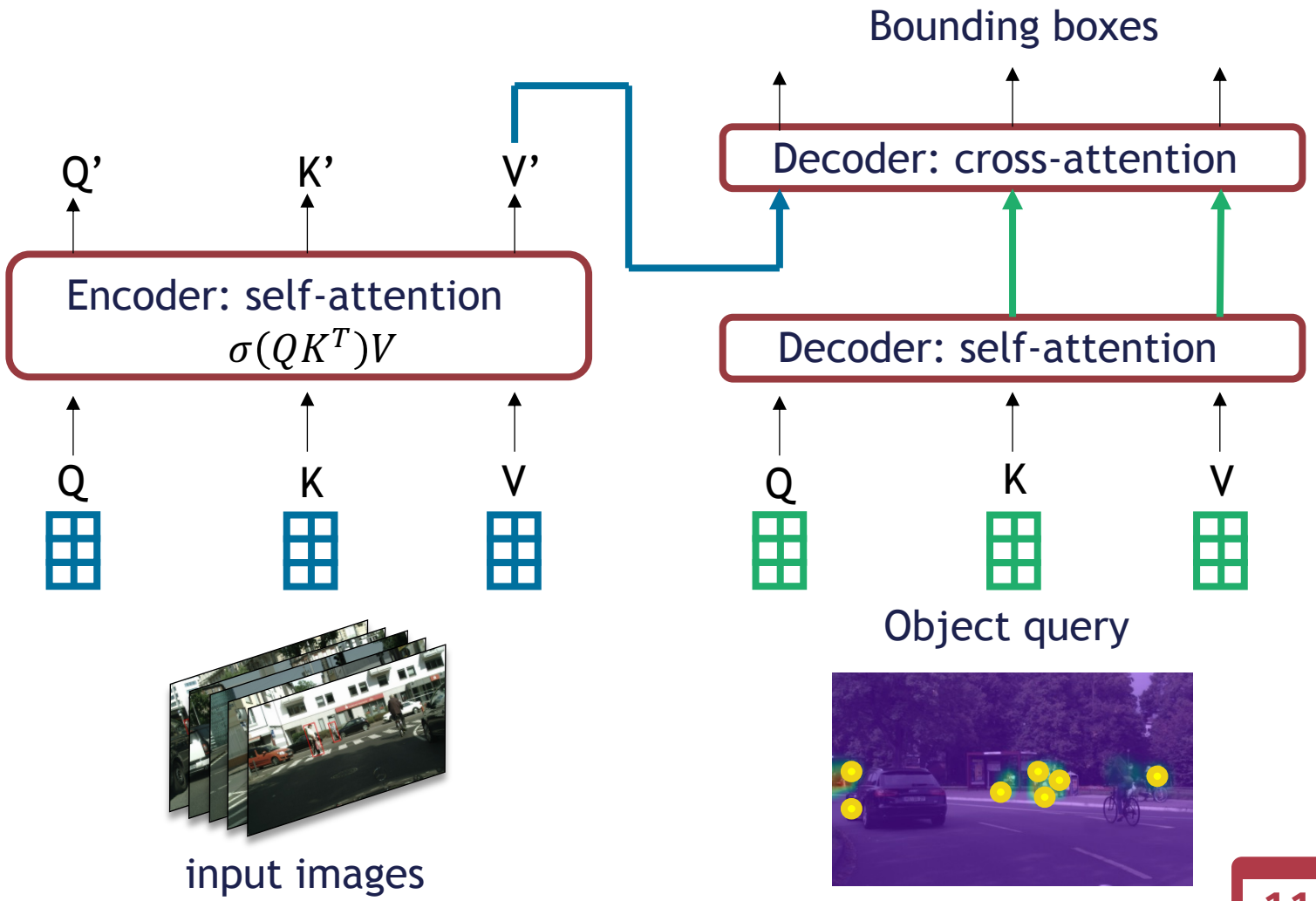
Knowledge-augmented Detection with Transformer

- Synthetic pedestrian or vehicle image patches support sequence of input images
- The same backbone is used to extract features from synthetic support image patches and input images
- Goal: Generate attention heatmaps between support images and input images. The heatmap contains knowledge of positions of objects of interest



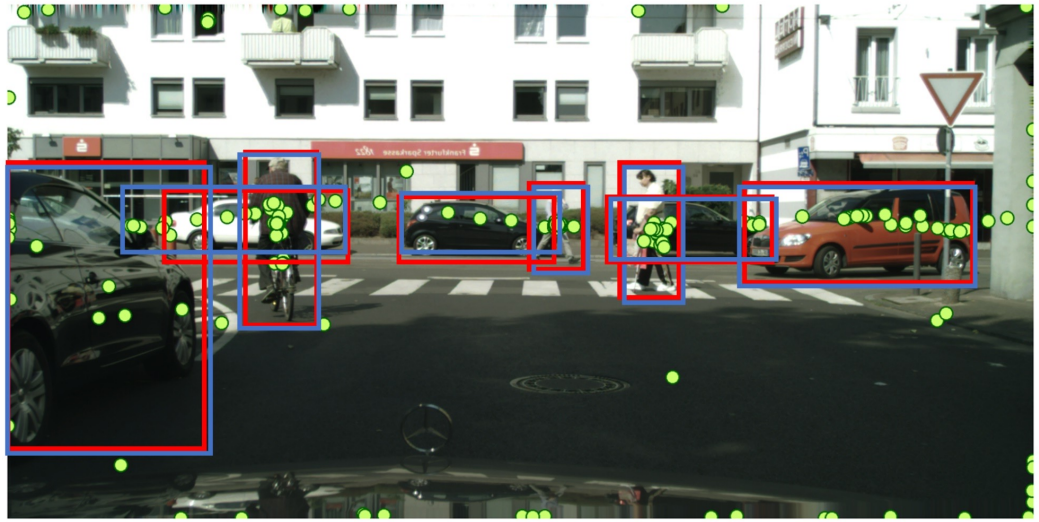
Knowledge-augmented Detection with Transformer

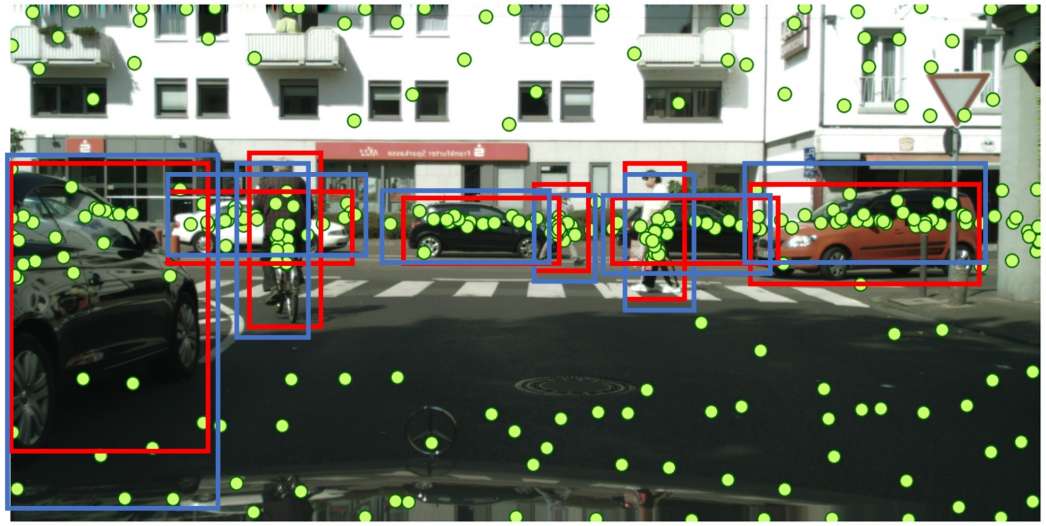
- Transformer-based detection uses object queries in decoder part for bounding box prediction.
- Instead of random initialized positional encoding in object queries, we add positional encoding based on attention heatmaps from encoder.



Knowledge-augmented Detection with Transformer

- Prior knowledge directs queries to focus more on targeting objects over baseline models. This results in increased accuracy in the positions of bounding boxes.

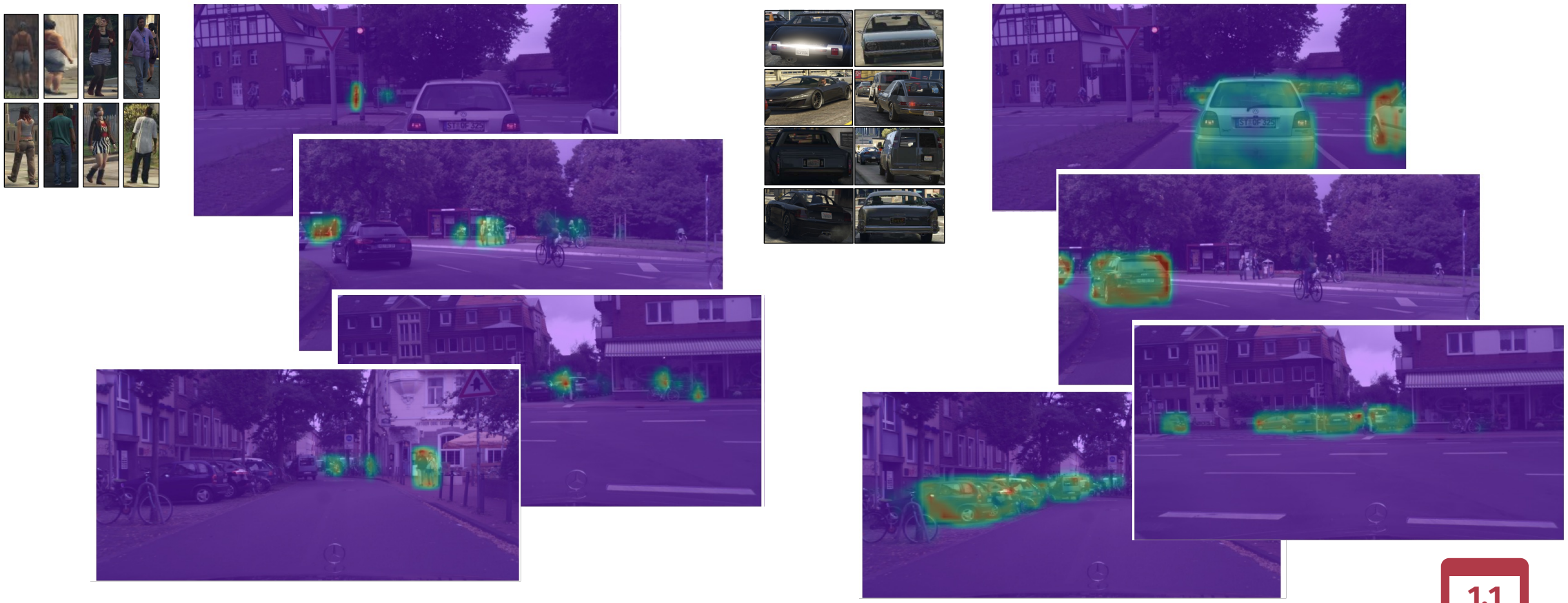




Left: our method; Right: baseline model
Red bounding box: ground-truth; Blue bounding box: predicted results

Knowledge-augmented Detection with Transformer

- With different support images, we can disentangle high-level semantics:



Knowledge-augmented Detection with Transformer

- Full Data Training: Utilizing the entire KITTI or Cityscapes dataset for training.
- Limited Data Training: Employing only a quarter of the original training data size.
- Despite observing a decrease in performance compared to training with the full dataset, we experience a mitigated decline.

	KITTI (AP)		Cityscapes (AP)	
	Full	Limited	Full	Limited
DETR	57.7%	52.0%	44.8%	40.1%
Ours	**66.3%**	**60.6%**	**51.4%**	**49.0%**

KI WISSEN

Automotive AI Powered by Knowledge

Julian Wörmann | fortiss | woermann@fortiss.org

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