Joint Prediction of Amodal and Visible Semantic Segmentation for Automated Driving

Jasmin Breitenstein,
Tim Fingscheidt | TU BS

Motivation

Amodal perception means to recognize the full shape of (partially) occluded objects. In automated driving this is crucial for safe environment perception. We will train amodal and visible semantic segmentation separately and jointly. We find that joint training can improve both tasks.

Proposed Joint Training Method

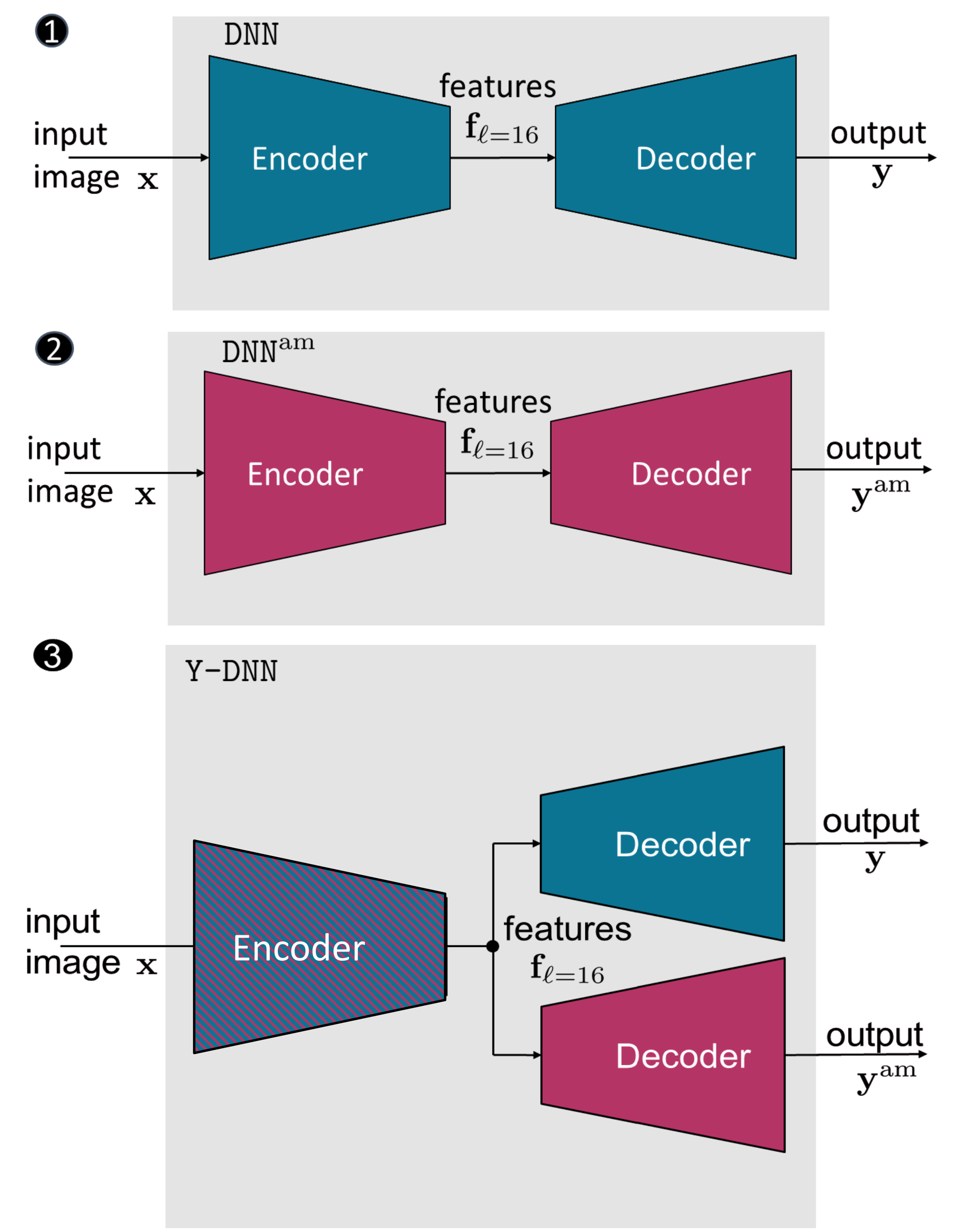

Figure 1: Training methods: ❶ only visible, ❷ only amodal, and ❸ joint training. Network choice: ERFNet

Amodal Cityscapes Challenge:

Amodal Cityscapes [1] training set $\mathcal{D}^{\mathrm{amCS}}_{\mathrm{train}}$
Amodal Cityscapes validation set $\mathcal{D}^{\mathrm{amCS}}_{\mathrm{val}}$
Evaluation on Cityscapes [2] validation set $\mathcal{D}^{\mathrm{CS}}_{\mathrm{val}}$ and Amodal Cityscapes test set $\mathcal{D}^{\mathrm{amCS}}_{\mathrm{test}}$
Evaluation metrics: mean intersection over union for visible pixels (mIoU) and mean intersection over union for invisible (occluded) pixels ($\mathrm{mIoU}^{\mathrm{inv}}$)

Experimental Results:

Method	$\mathcal{D}^{\mathrm{amCS}}_{\mathrm{test}}$		$\mathcal{D}^{\mathrm{CS}}_{\mathrm{val}}$	
	mIoU	$\mathrm{mIoU}^{\mathrm{inv}}$	mIoU	$\mathrm{mIoU}^{\mathrm{inv}}$
ERFNet	62.99%	5.00%	67.21%	*
$\mathrm{ERFNet}^{\mathrm{am}}$	20.16%	36.48%	21.00%	*
Y-ERFNet (ours)	**63.32%**	**43.32%**	**68.35%**	*

*Table 1: Performance of ERFNet, $\mathrm{ERFNet}^{\mathrm{am}}$, and Y-ERFNet on $\mathcal{D}^{\mathrm{amCS}}_{\mathrm{test}}$ and $\mathcal{D}^{\mathrm{CS}}_{\mathrm{val}}$ (training on $\mathcal{D}^{\mathrm{amCS}}_{\mathrm{train}}$ for 120 epochs). Best results are shown in **bold**. * cannot be calculated due to missing ground truth.*

Especially for vulnerable road users (person, rider) the joint training method improves the $\mathrm{mIoU}^{\mathrm{inv}}$. Detailed results for per-class $\mathrm{IoU}^{\mathrm{inv}}$ and IoU are in the paper [3].

Qualitative Results of the Joint Training

Figure 2 shows the qualitative predictions for mobile objects. Mobile objects are all objects with the ability to move, e.g., person, rider, car. We especially visualize the relevant parts, i.e., the insertion of the amodal prediction into the predicted mobile objects. We see that the occluded person behind the car is anticipated by the joint training and the full person shape is recovered.

Conclusions:

All methods allow the prediction of separate classes (as opposed to groupwise settings). Single-task amodal semantic segmentation improves amodal performance. Joint training improves both amodal and visible semantic segmentation.

References:

[1] J. Breitenstein, T. Fingscheidt, "Amodal Cityscapes: A New Dataset, its Generation, and an Amodal Semantic Segmentation Challenge Baseline," in Proc. of IV 2022

[2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, B. Schiele, "The Cityscapes Dataset for Semantic Urban Scene Understanding," in Proc. of CVPR 2016

[3] J. Breitenstein, J. Löhdefink, T. Fingscheidt, "Joint Prediction of Amodal and Visible Semantic Segmentation for Automated Driving," in Proc. of ECCV-Workshops 2022

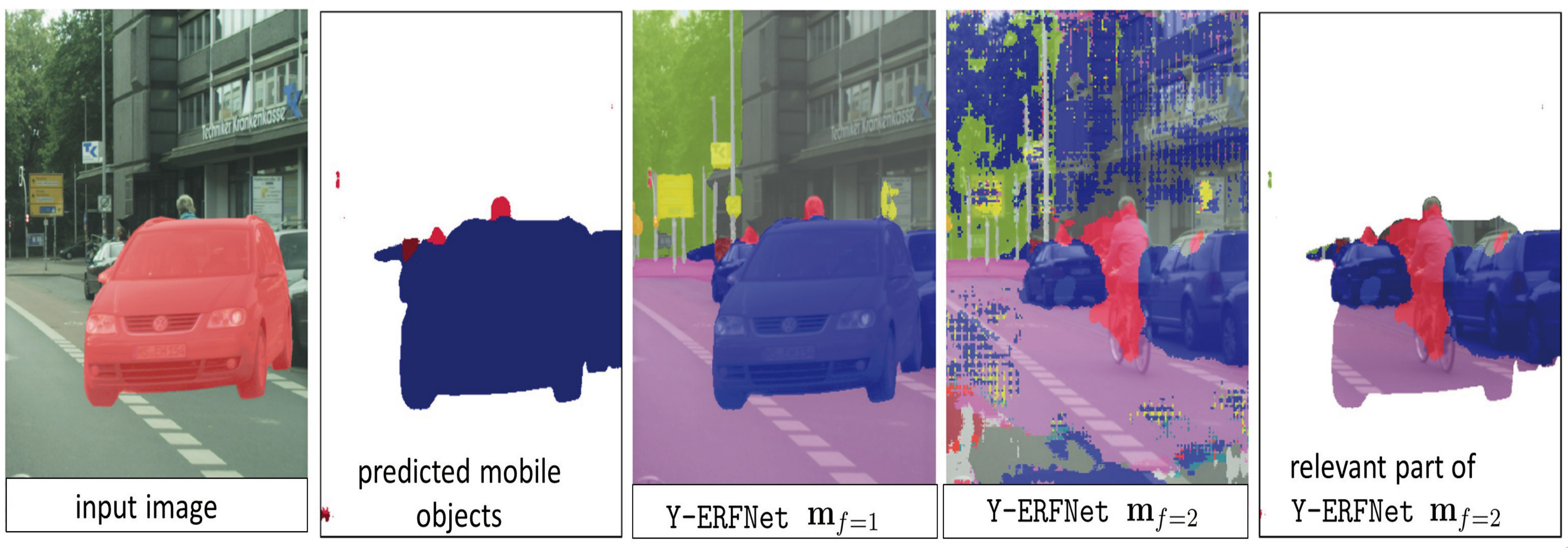

Figure 2: Qualitative results of the joint training method showing that the full shape of the person occluded by the car can be recovered. The rider cannot be anticipated from just the head.

ML Function Development

Partners

External partners

For more information contact:
j.breitenstein@tu-bs.de

KI Data Tooling 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.ki-datatooling.de   X @KI_Familie   KI Familie

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