Supplementary Material: Discovering Latent Classes for Semi-Supervised Semantic Segmentation

Olga Zatsarynna[1⋆], Johann Sawatzky[1,2⋆], and Juergen Gall[1]

[1] University of Bonn
[2] EyewareTech
{s6olzats, jsawatzk, jgall} @ uni-bonn.de

1 Training Details

The optimization of the segmentation network is performed using SGD with a momentum equal to 0.9 and the learning rate decay of 10^{-4}. The learning rate, that is initially equal to $2.5 \cdot 10^{-4}$, is decreased with polynomial decay with the power of 0.9. For the discriminator, we employ the Adam optimizer [2], where the initial learning rate is equal to 10^{-4} and that follows the same decay schedule as introduced for the segmentation network.

At each iteration, we alternately apply the described training scheme on the batch of the randomly sampled labeled and unlabeled data. To ensure the robustness of the evaluation procedure, we report results averaged over 5 random seeds that control the sampling procedure. We add the consistency loss term only after 5000 iterations since the latent branch needs to learn some useful latent classes first.

On Pascal VOC 2012, during the training procedure, the images are cropped with crop size equal to 321×321 and undergo random scaling and horizontal mirroring. We train our model for 20k iterations with a batch size of 10 images. The testing of the resulting model is carried out on the validation set.

On the Cityscapes dataset, during training, we pre-process the images by performing cropping operations with crop size equal to 505×505 and additionally apply random scaling and horizontal mirroring. On the Cityscapes dataset, our model is trained for 40k iterations with batches of size 2. We report the results of testing the resulting model on the validation set.

2 IIT Affordances

The IIT Affordances dataset [3] contains images of 10 common human tools. It has 8835 images in total, where 50% are used for the training split, 20% for the validation split, and the rest 30% for the test split. Around 60% of the images in the dataset are from ImageNet, while the rest are taken from cluttered scenes, which implies a large variation of images within the dataset.

⋆ contributed equally

Table 1. Comparison to Hung et.al on IIT Affordances. We used 7 latent classes for the proposed model

IIT 2017 Affordances			
	Fraction of annotated images		
Method	1/50	1/20	1/8
	mIoU (%)		
Hung et al. [1]	47.4	55.8	64.3
Proposed	**51.3**	**58.8**	**65.4**

During training, the images are cropped with the crop size equal to 321×321 and undergo random scaling and horizontal mirroring. We train our model for 20k iterations with a batch size of 10 images on the training and validation images together. The testing of the resulting model is carried out on the test set. We report the results in the Table 1. As for the other datasets, our approach outperforms [1].

3 Qualitative Results

Figures 1, 2 and 3 show additional qualitative results on Pascal VOC 2012, Cityscapes and IIT Affordances, respectively.

4 Manual assignment

Table 2 lists the manual assignment of the semantic classes to 10 supercategories as it is used for the experiment manual in Table 5 of the paper.

Table 2. Manual assignment of Pascal VOC 2012 classes to 10 supercategories that we use instead of learned latent classes in the ablation study.

Mapping of semantic classes to supercategories	
Manually defined supercategory	VOC semantic classes
Background	Background
Aeroplane	Aeroplane
Bicycle	Bicycle
Bird	Bird
Boat	Boat
Person	Person
Ground vehicle with engine	Bus, car, motorbike, train
Mammal	Cat, cow, dog, horse, sheep
Furniture	Dinning table, sofa, chair
Miscellaneous	Bottle, tv monitor, potted plant

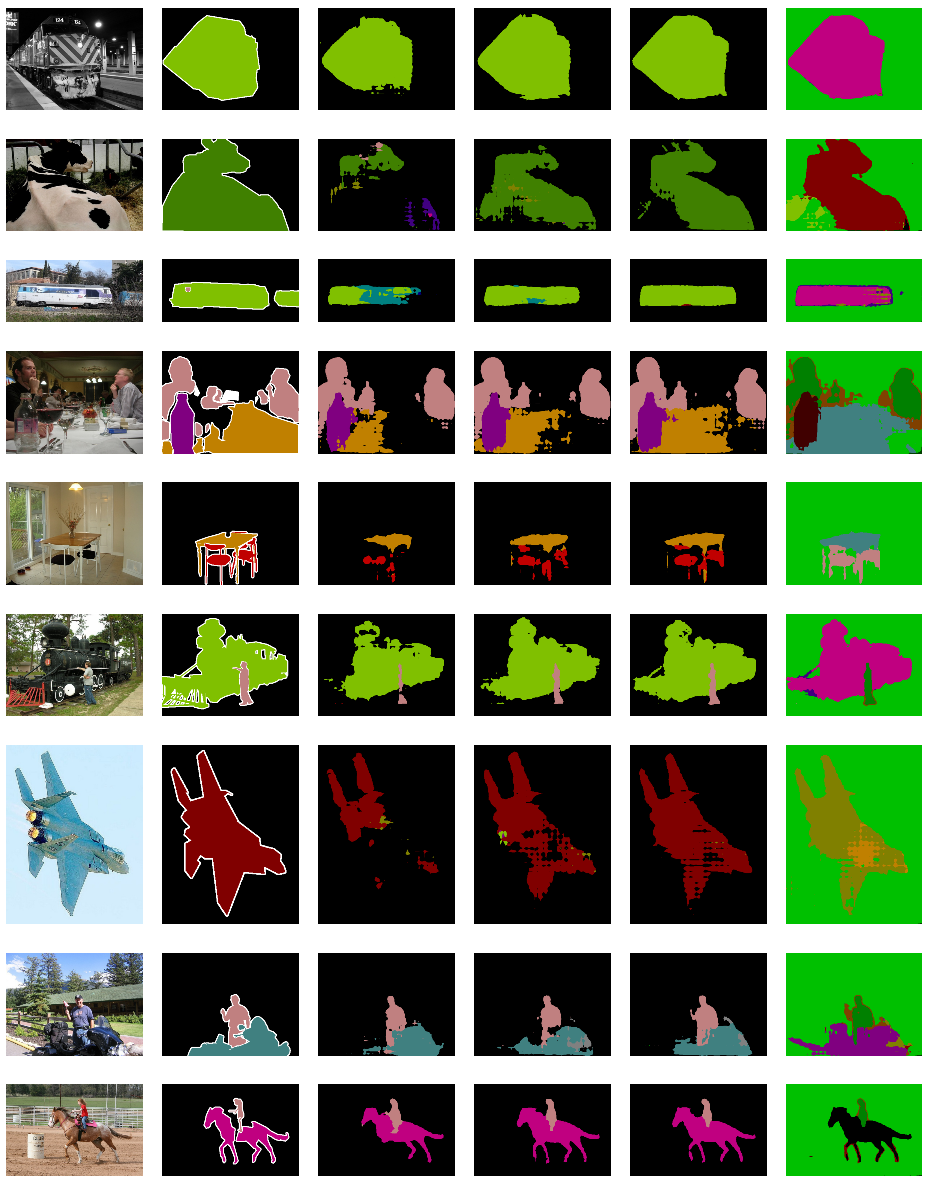

Fig. 1. Qualitative examples from the Pascal VOC 2012 val set. From left to right: image, ground truth, L_{ce}, proposed without adversarial loss, proposed, latent classes.

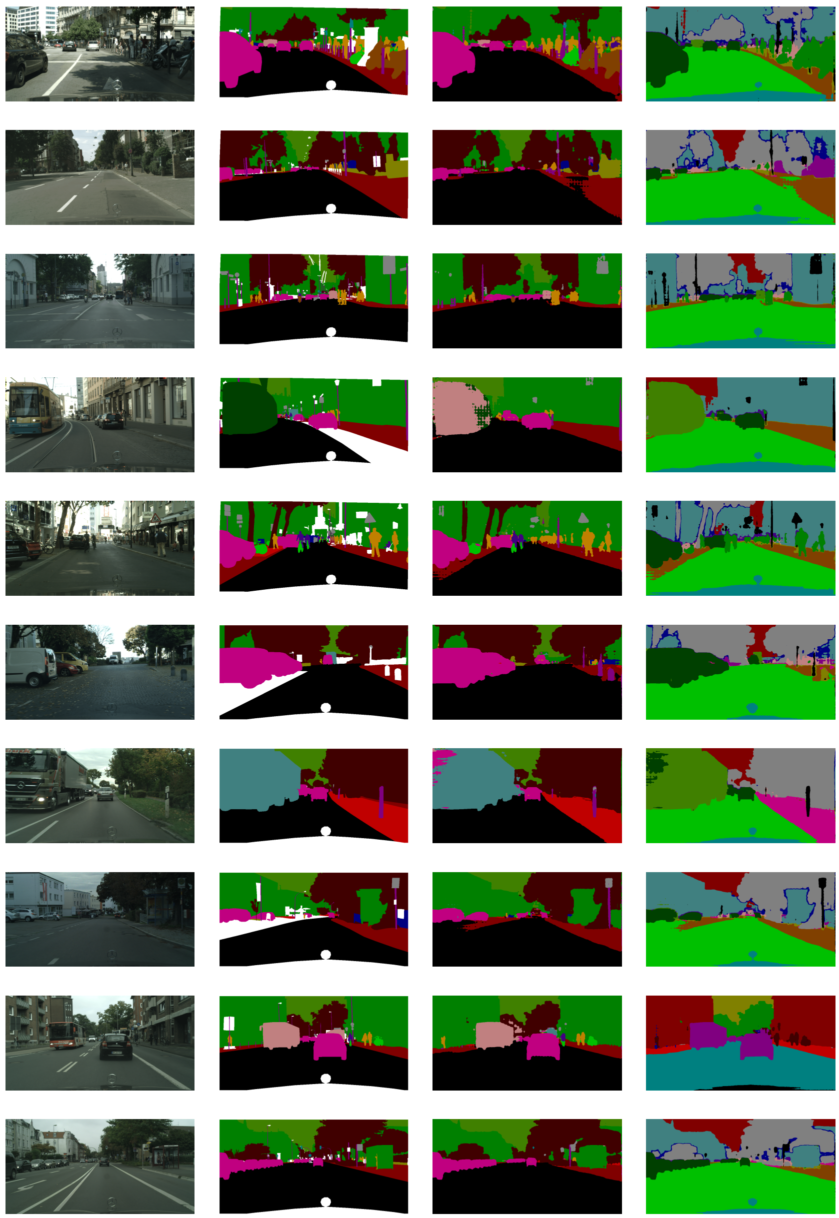

Fig. 2. Qualitative examples from the Cityscapes val set. From left to right: image, ground truth, proposed, latent classes.

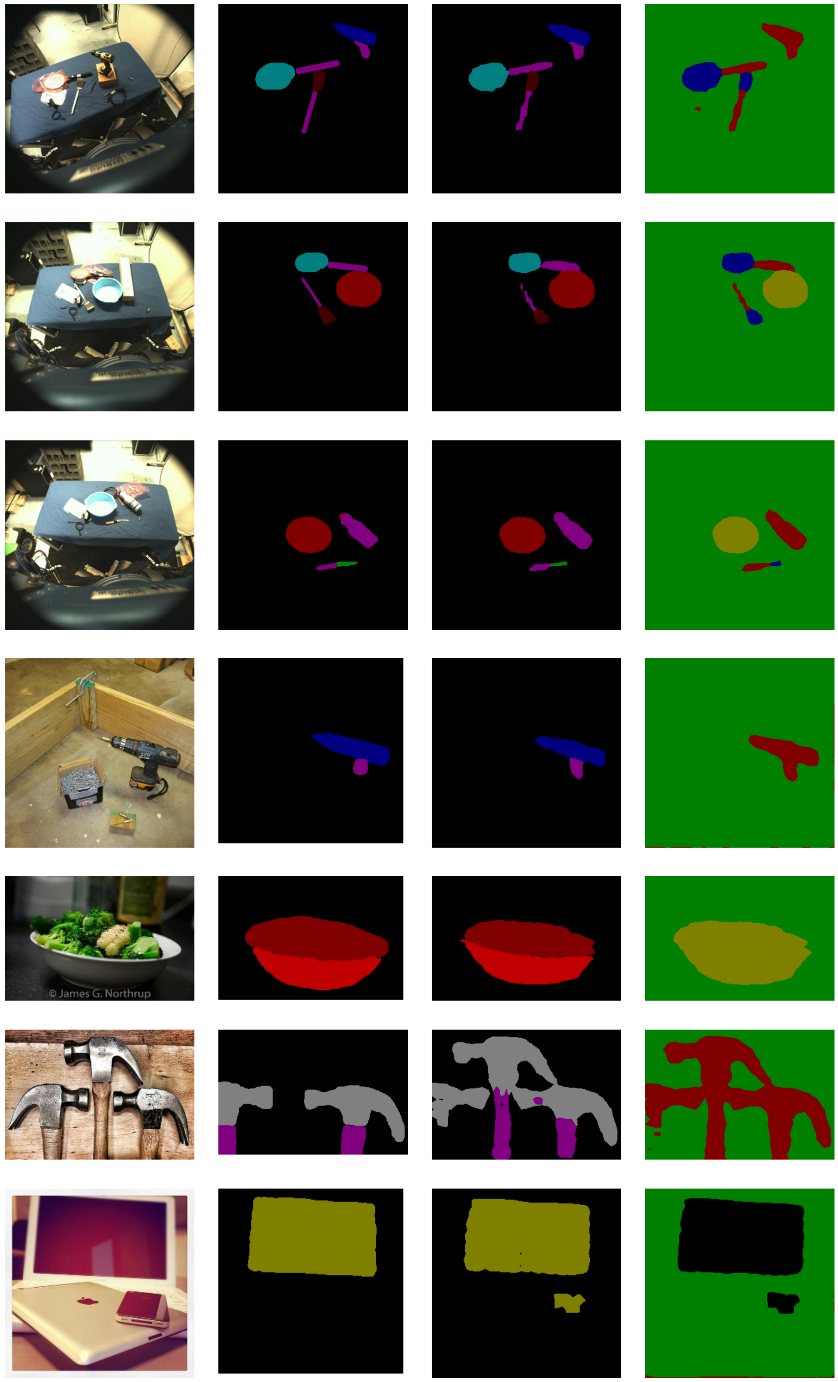

Fig. 3. Qualitative examples from the IIT Affordances test set. From left to right: image, ground truth, proposed, latent classes.

References

1. Hung, W.C., Tsai, Y.H., Liou, Y.T., Lin, Y.Y., Yang, M.H.: Adversarial learning for semi-supervised semantic segmentation. In: Proceedings of the British Machine Vision Conference (BMVC) (2018) 2
2. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. ArXiv **abs/1412.6980** (2014) 1
3. Nguyen, A., Kanoulas, D., Caldwell, D.G., Tsagarakis, N.: Object-based affordances detection with convolutional neural networks and dense conditional random fields. In: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (2017) 1