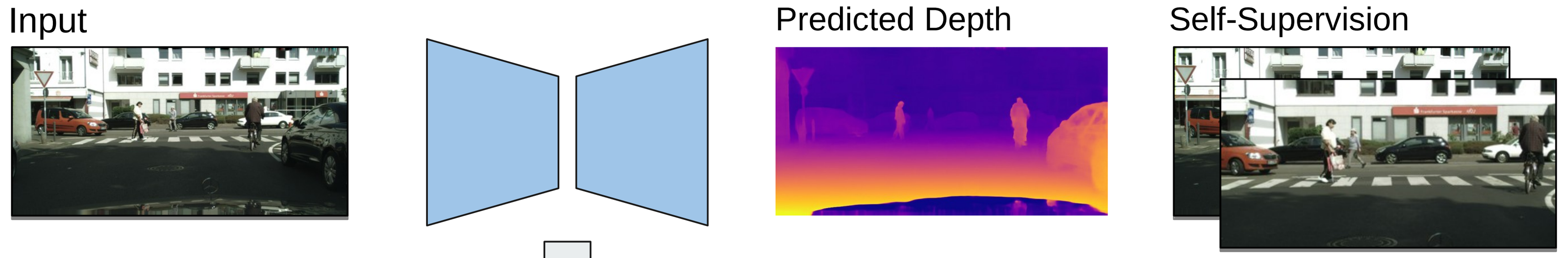
Three Ways to Improve Semantic Segmentation with Self-Supervised Depth Estimation

Lukas Hoyer[1], Dengxin Dai[1], Yuhua Chen[1], Adrian Köring[2], Suman Saha[1], Luc Van Gool[1,3]
[1] ETH Zürich, Switzerland [2] University of Bonn, Germany [3] KU Leuven, Belgium

1 Concept

Self-Supervised Depth Estimation (SDE)

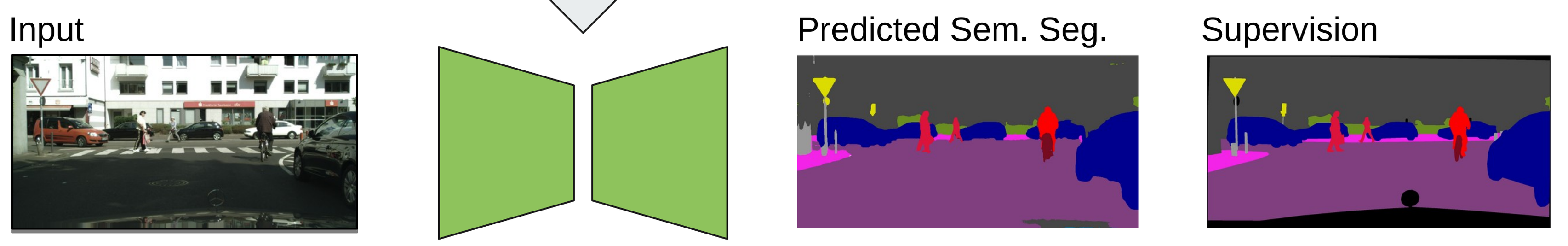
+ Requires only unlabeled image sequences



Semantic Segmentation

- Requires expensive human annotation

Knowledge Transfer
- Transfer- and Multi-Task Learning
- DepthMix Data Augmentation
- Automatic Data Selection

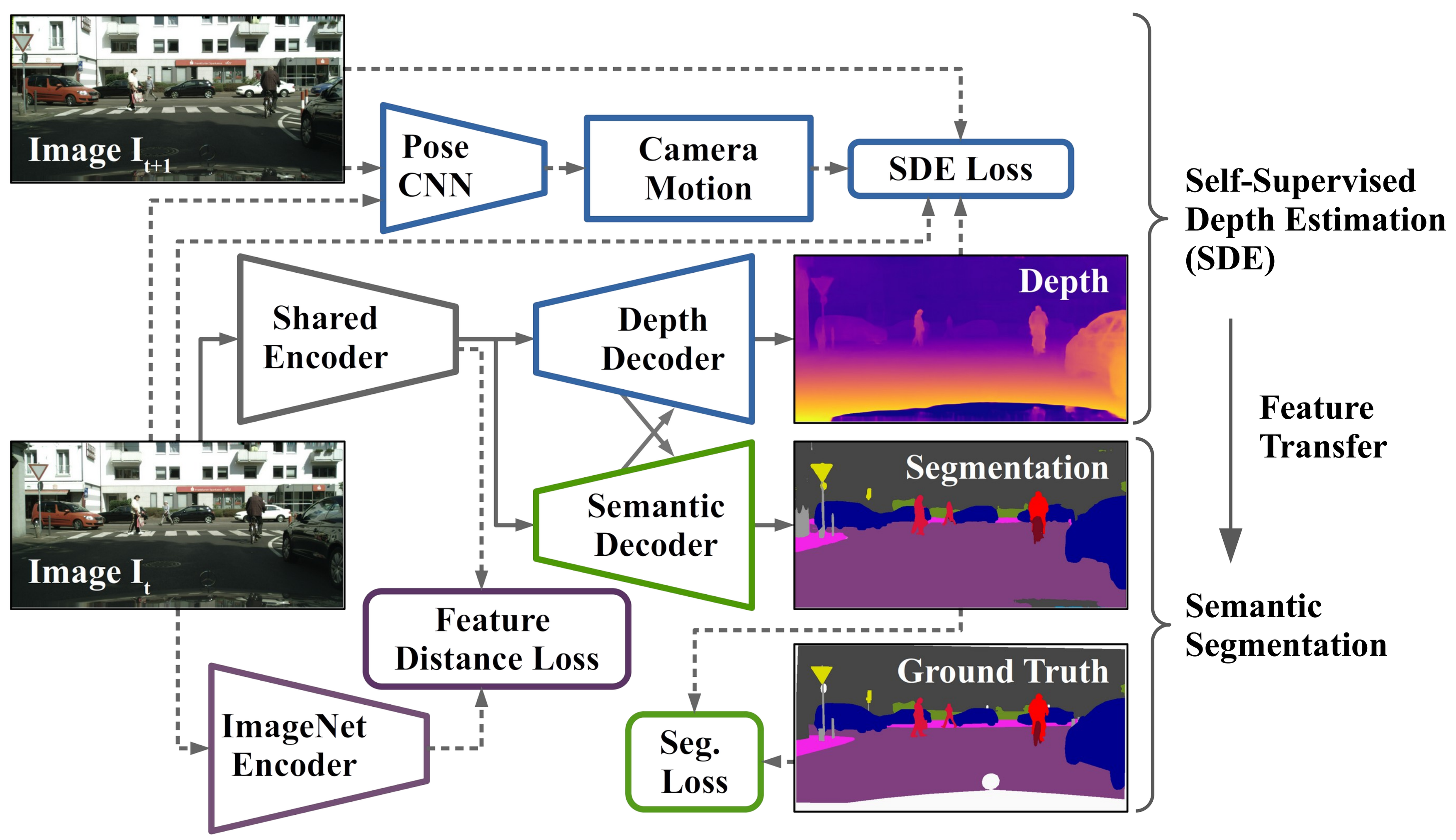


Advantage: Knowledge from SDE reduces necessary annotations for semantic segm.
Implementation: https://github.com/lhoyer/improving_segmentation_with_selfsupervised_depth

2 Transfer and Multi-Task Learning

Motivation: Utilize depth features for semantic segmentation
Approach:
- Transfer Learning: Initialize segmentation branch with depth pretraining
- Multi-Task Learning: Exchange features between depth and segmentation decoder

Advantage: SDE learns features from a large set of unlabeled image sequences



3 DepthMix

Concept

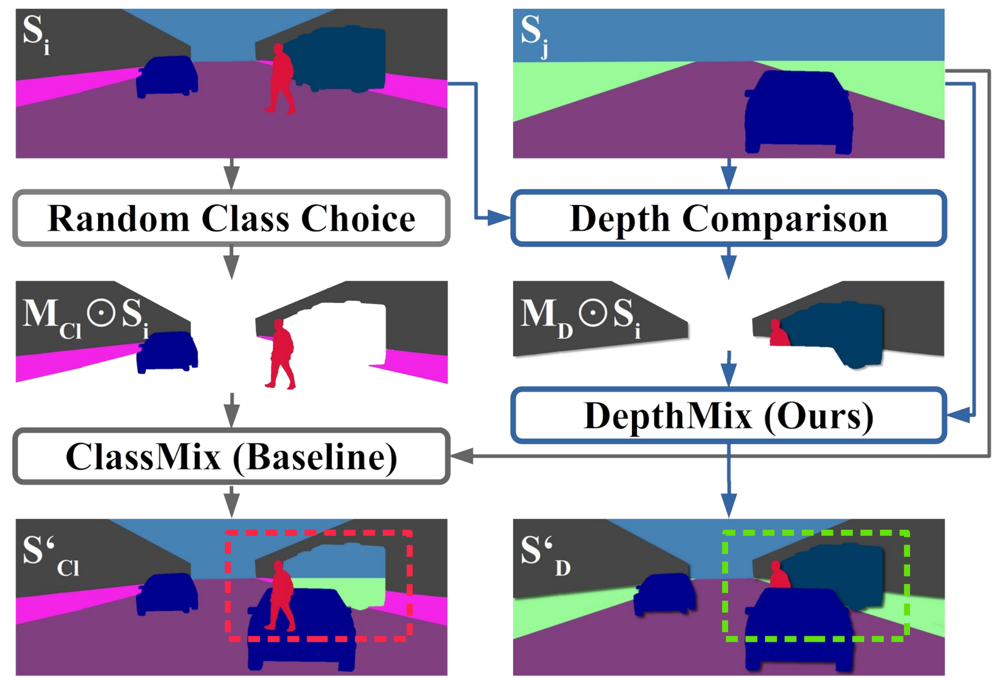
Motivation: Mitigate occlusion artifacts
Approach: Select pixels closer to camera

Mix Mask:
$$M(a,b) = \begin{cases} 1 & \text{if } \hat{D}_i(a,b) < \hat{D}_j(a,b) + \epsilon \\ 0 & \text{otherwise} \end{cases}$$

Mixed Image:
$$I' = M \odot I_i + (1-M) \odot I_j$$

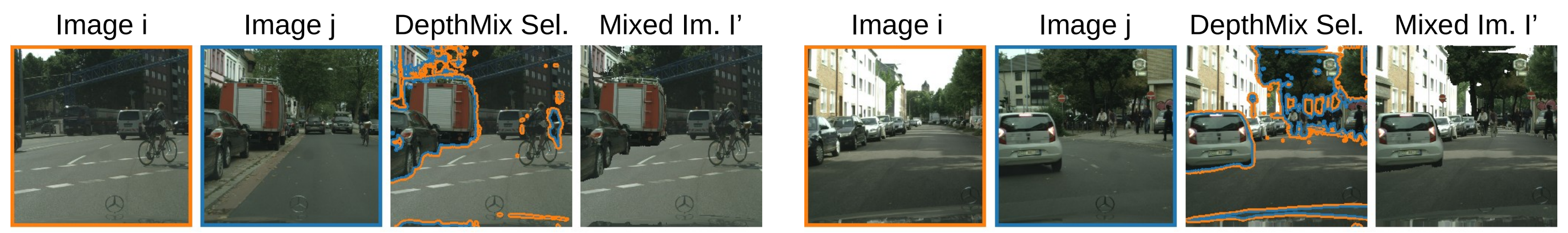
Mixed Segmentation:
$$S' = M \odot S_i + (1-M) \odot S_j$$

Advantage: Geometrically valid mixing



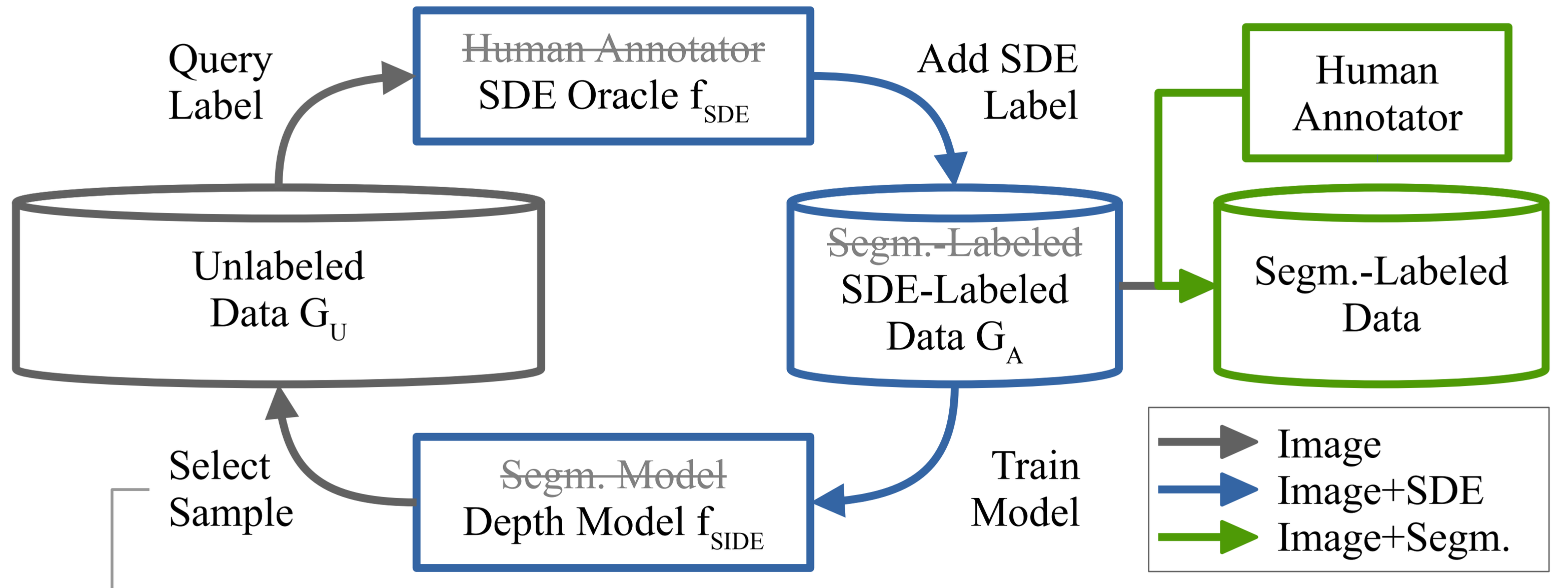
Real-World Examples

DepthMix produces precise mixing boundaries and effectively handles occlusions



4 Automatic Data Selection for Annotation

Motivation: Select most beneficial samples to be annotated
Approach: Iterative sample selection with SDE proxy-task oracle
Advantage: No human in the loop → increased flexiblity, efficiency, and scalability



- Diversity Sampling: Select sample i with high SDE feature Φ^{SDE} distance to G_A

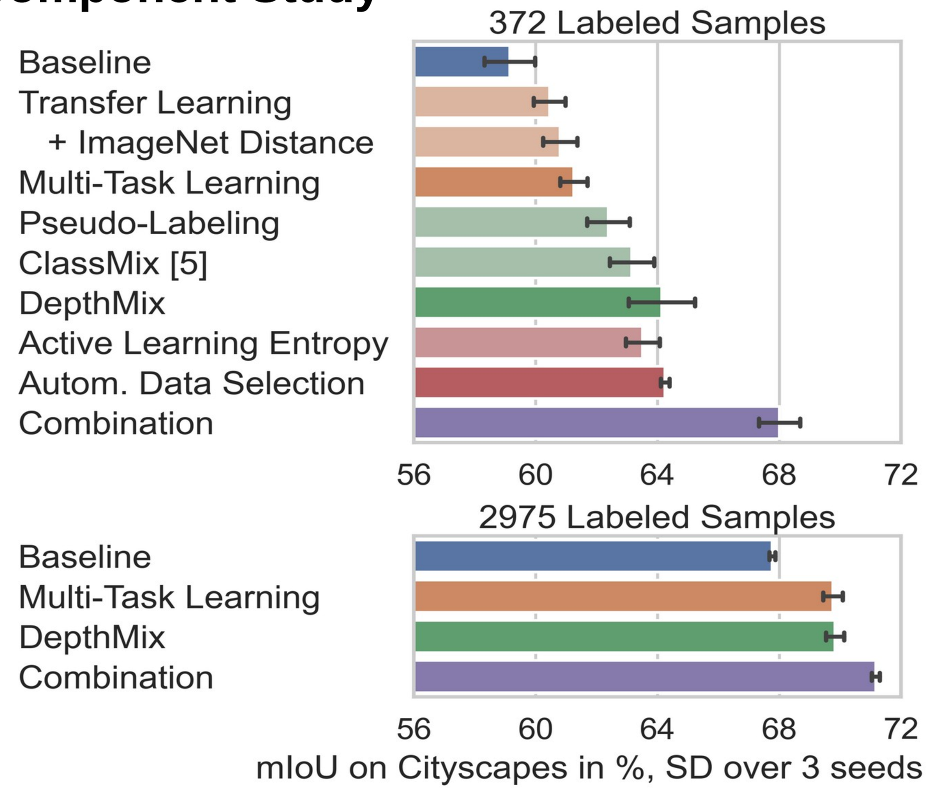
$$i = \operatorname*{argmax}_{I_i \in G_U} \min_{I_j \in G_A} \|\Phi_i^{SDE} - \Phi_j^{SDE}\|_2$$

- Uncertainty Sampling: Select sample i with high depth student error

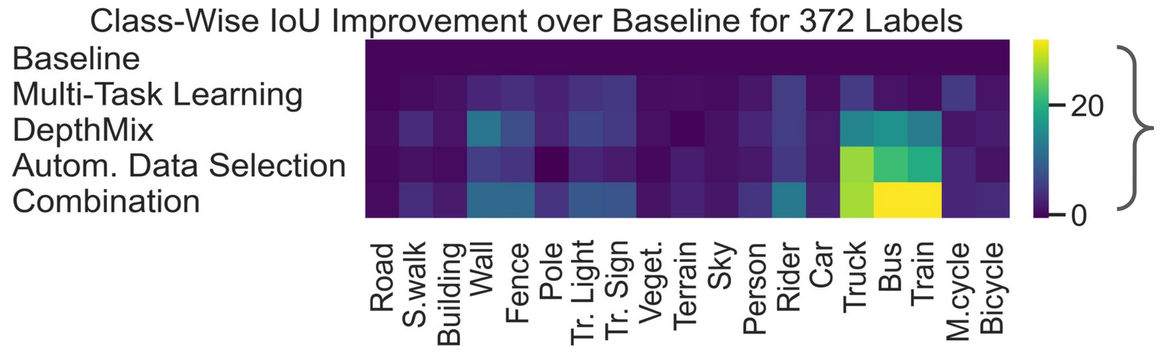
$$i = \operatorname*{argmax}_{I_i \in G_U} \|\log(1 + f_{SDE}(I_i)) - \log(1 + f_{SIDE}(I_i))\|_1$$

5 Evaluation on Cityscapes

Component Study



Each contribution significantly outperforms baseline
- Transfer and Multi-Task Learning enable effective feature transfer
- Mixing improves pseudo-labeling
- DepthMix outperforms ClassMix [5]
- Autom. Data Selection outperforms human-in-the-loop active learning with entropy
- Combination further improves performance
- Combination achieves fully-supervised baseline performance with only 1/8 labels
- Our contributions can also improve the fully-supervised performance



- Multi-Task Learning improves classes with depth discontinuities
- Autom. Data Selection improves difficult classes
- DepthMix improves both

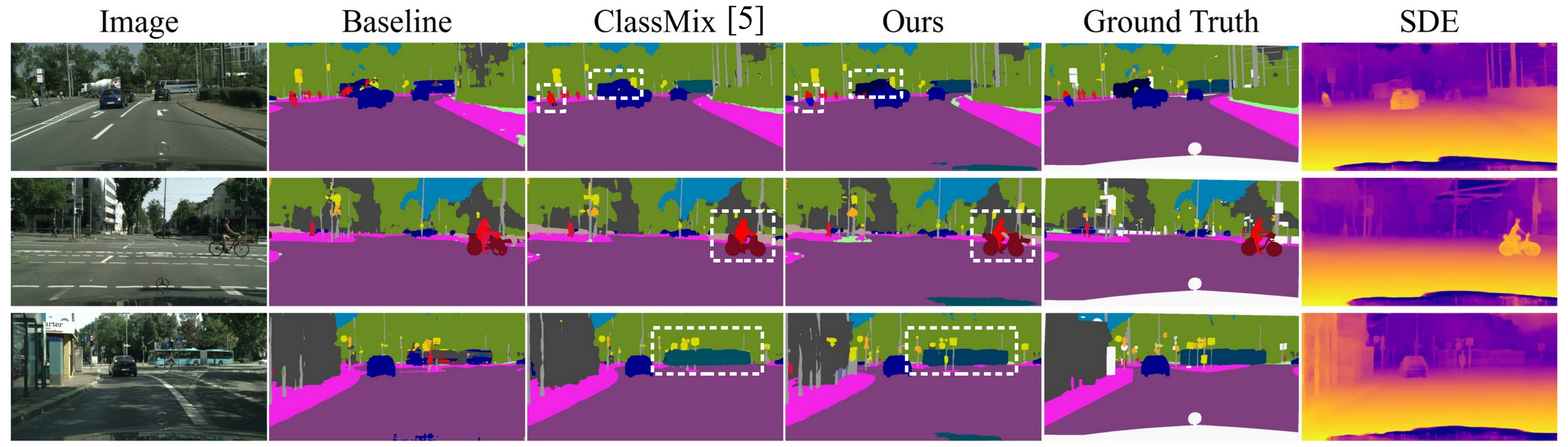
Comparison with State of the Art

Absolute mIoU Improvement over Baseline

Labeled Samples	100	372	744	2975
Adversarial [3]		+3.30	+2.40	
s4GAN [4]		+3.10	+1.70	-0.20
DST-CBC [1]	+3.20	+3.80	+3.30	
CutMix [2]	+6.79	+5.09	+3.30	+0.15
ClassMix [5]	+10.23	+6.51	+3.55	
Ours (Combination)	+13.34	+8.87	+5.92	+3.40

- Our method outperforms previous SotA methods by a significant margin

Example Predictions



- Ours better distinguishes difficult classes (e.g. truck, train, and bus)
- Ours segments finer structures at depth discontinuities (e.g. rider, pole, and tr. sign)

trained on 100 labeled samples

6 References

[1] Feng et al. "Semi-supervised semantic segmentation via dynamic self-training and classbalanced curriculum." arXiv preprint. 2020.
[2] French et al. "Semi-supervised semantic segmentation needs strong, varied perturbations." BMVC. 2020.
[3] Hung et al. "Adversarial learning for semi-supervised semantic segmentation." BMVC. 2019.
[4] Mittal et al. "Semi-supervised semantic segmentation with high-and low-level consistency." PAMI, 2019.
[5] Olsson et al. "Classmix: Segmentation-based data augmentation for semi-supervised learning." WACV. 2021.