

SEMANTICALLY-ADAPTIVE UPSAMPLING FOR LAYOUT-TO-IMAGE TRANSLATION –SUPPLEMENTARY DOCUMENT–

Anonymous authors

Paper under double-blind review

We are providing additional experimental results on the layout-to-image translation task. First, we compare the proposed method with two leading methods, i.e., GauGAN (Park et al., 2019) and CC-FPSE (Liu et al., 2019). Additionally, we provide the visualization results of the generated semantic maps.

1 COMPARISONS WITH STATE-OF-THE-ART

In this section, we show more generation results of the proposed method compared with those from two leading layout-to-image translation models, i.e., GauGAN (Park et al., 2019) and CC-FPSE (Liu et al., 2019). The results of CelebAMask-HQ (Lee et al., 2020), Facades (Tyleček & Šára, 2013), and DeepFashion (Liu et al., 2016) are shown in Fig. 1, 3, 2, 4, 5, and 6.

The results of Cityscapes (Cordts et al., 2016), ADE20K (Zhou et al., 2017), and COCO-Stuff (Caesar et al., 2018) are shown in Fig. 7, 8, 9, 10, 11 and 12. We observe that the model with the proposed SA-UpSample achieves visually better results than both competing methods on all the six datasets.

2 VISUALIZATION OF GENERATED SEMANTIC MAPS

We follow GauGAN (Park et al., 2019) and use DRN-D-105 (Yu et al., 2017) on the generated Cityscapes images to produce corresponding semantic maps. The results compared with those produced by GauGAN are shown in Fig. 13. We clearly see that the proposed method generates more semantically-consistent and realistic results than the original GauGAN model, which further validates the effectiveness of the proposed SA-UpSample.

REFERENCES

- Holger Caesar, Jasper Uijlings, and Vittorio Ferrari. Coco-stuff: Thing and stuff classes in context. In *CVPR*, 2018. 1
- Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016. 1
- Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. Maskgan: Towards diverse and interactive facial image manipulation. In *CVPR*, 2020. 1
- Xihui Liu, Guojun Yin, Jing Shao, Xiaogang Wang, et al. Learning to predict layout-to-image conditional convolutions for semantic image synthesis. In *NeurIPS*, 2019. 1, 9, 10, 11, 12, 13, 14
- Ziwei Liu, Ping Luo, Shi Qiu, Xiaogang Wang, and Xiaoou Tang. Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In *CVPR*, 2016. 1
- Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *CVPR*, 2019. 1, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15
- Radim Tyleček and Radim Šára. Spatial pattern templates for recognition of objects with regular structure. In *GCPR*, 2013. 1

Fisher Yu, Vladlen Koltun, and Thomas Funkhouser. Dilated residual networks. In *CVPR*, 2017. [1](#)

Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. Scene parsing through ade20k dataset. In *CVPR*, 2017. [1](#)



Figure 1: Qualitative comparison on CelebAMask-HQ. From left to right: Input, GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.



Figure 2: Qualitative comparison on CelebAMask-HQ. From left to right: Input, GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.



Figure 3: Qualitative comparison on CelebAMask-HQ. From left to right: Input, GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.



Figure 4: Qualitative comparison on Facades. From left to right: Input, GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

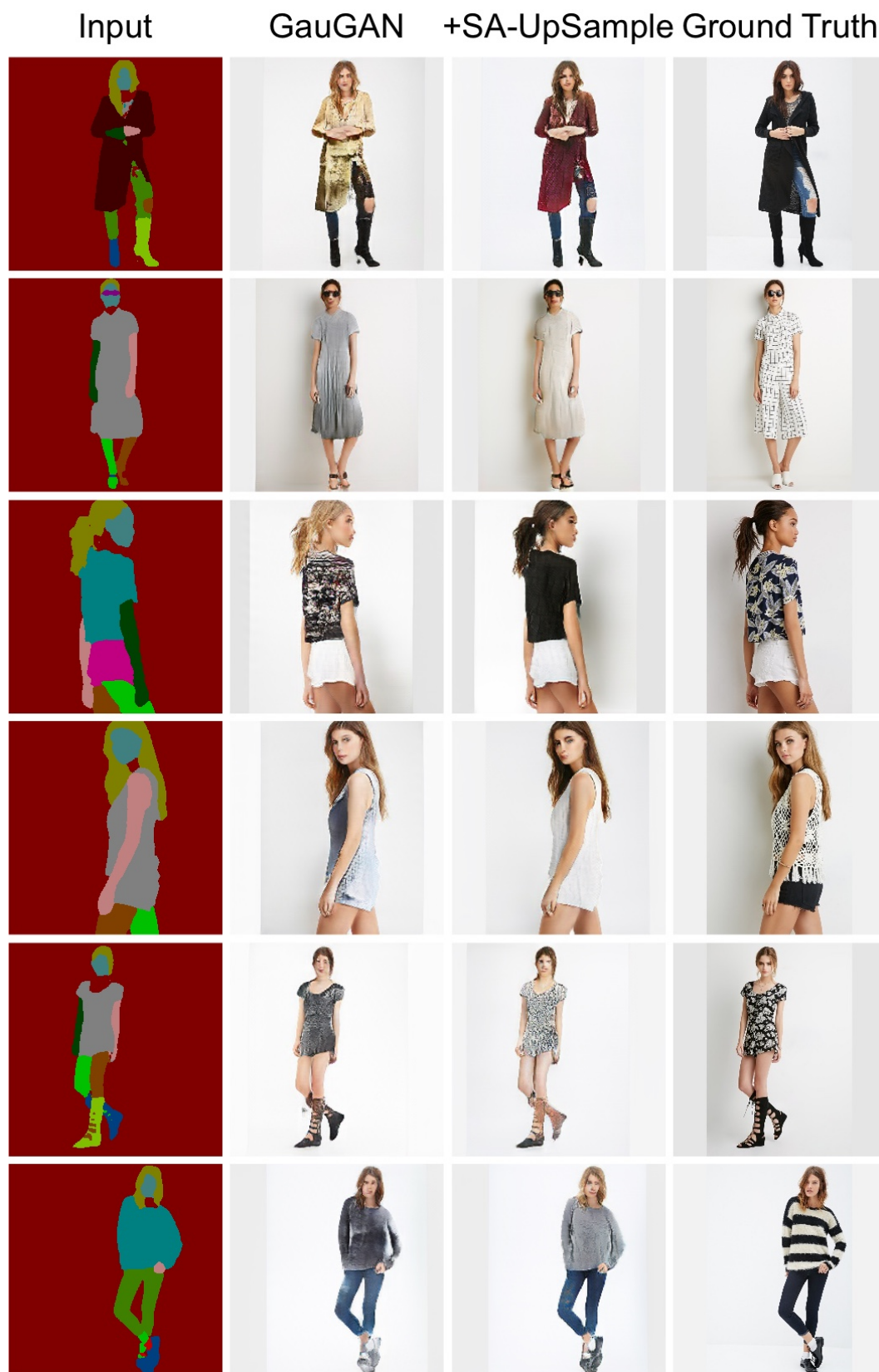


Figure 5: Qualitative comparison on DeepFashion. From left to right: Input, GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

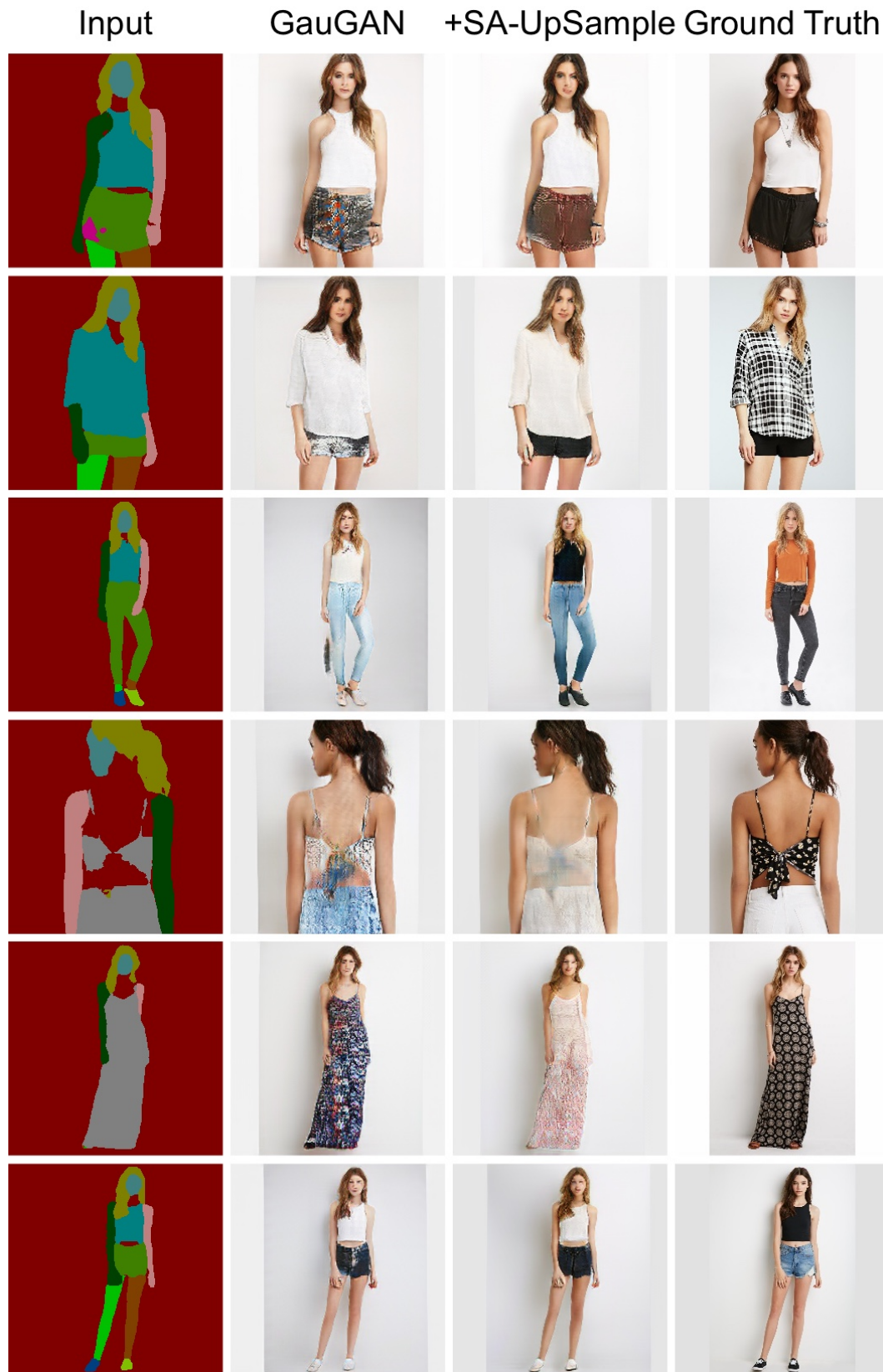




Figure 7: Qualitative comparison on Cityscapes. From left to right: Input, CC-FPSE (Liu et al., 2019), GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

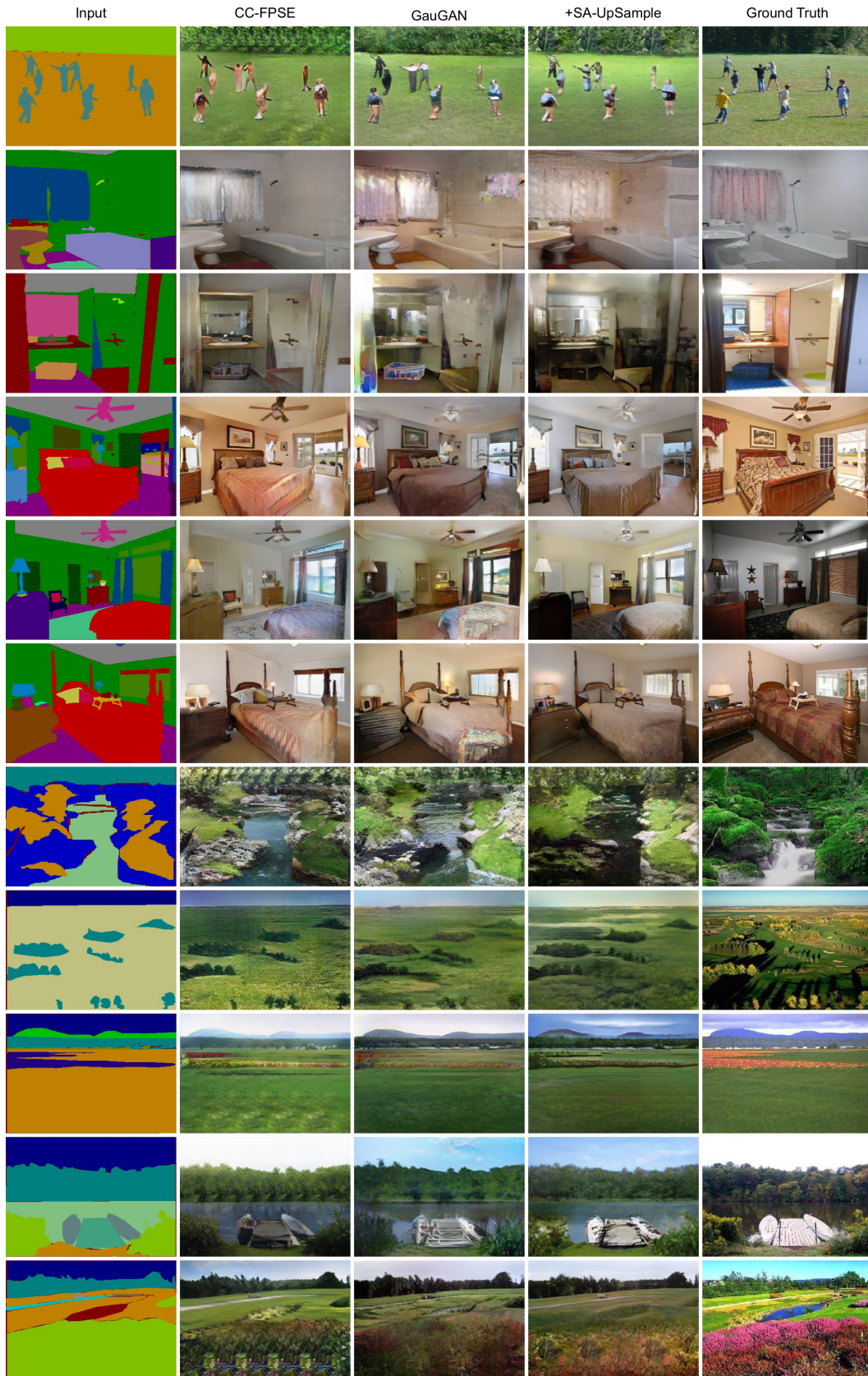


Figure 8: Qualitative comparison on ADE20K. From left to right: Input, CC-FPSE (Liu et al., 2019), GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

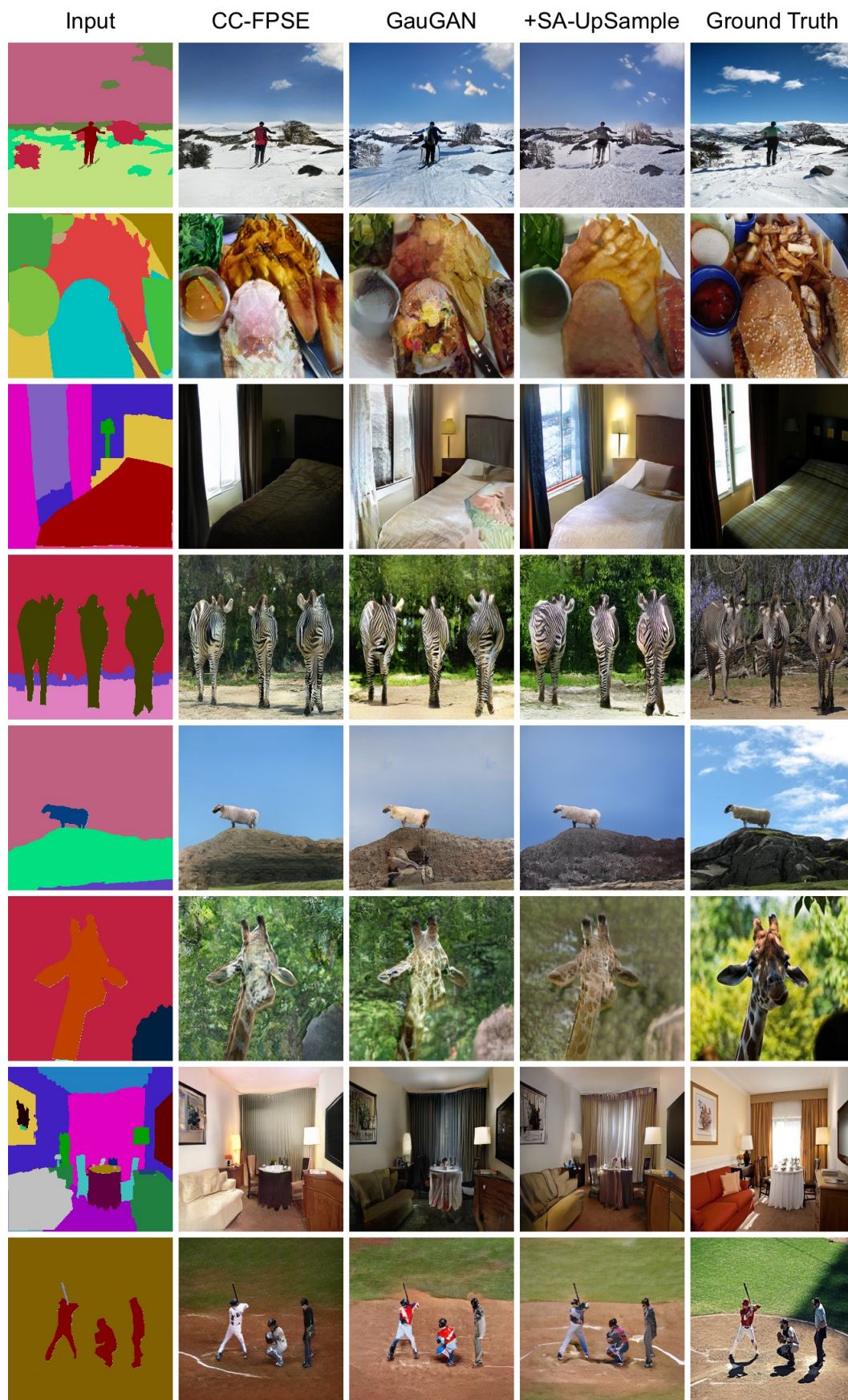


Figure 9: Qualitative comparison on COCO-Stuff. From left to right: Input, CC-FPSE (Liu et al., 2019), GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

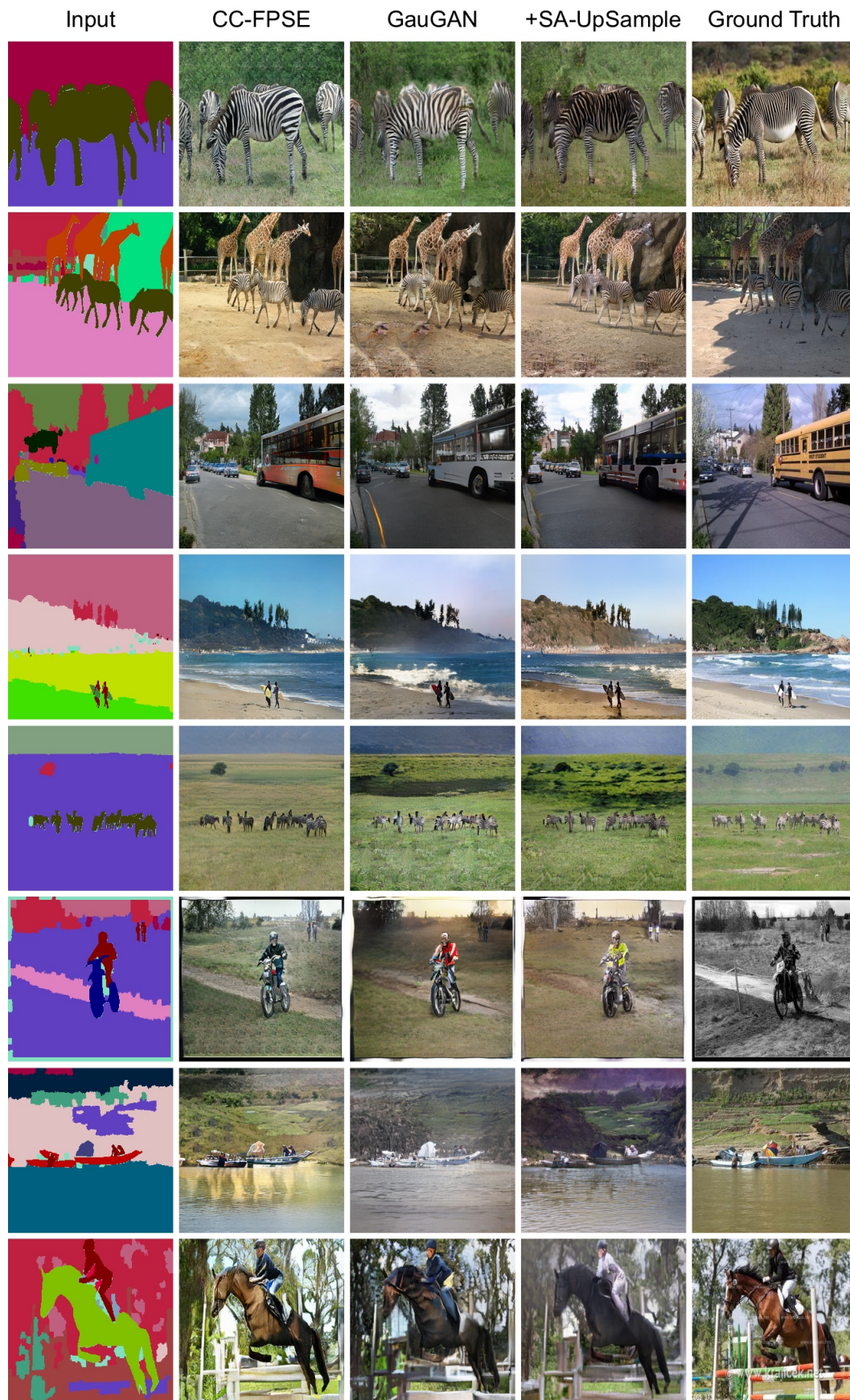


Figure 10: Qualitative comparison on COCO-Stuff. From left to right: Input, CC-FPSE (Liu et al., 2019), GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

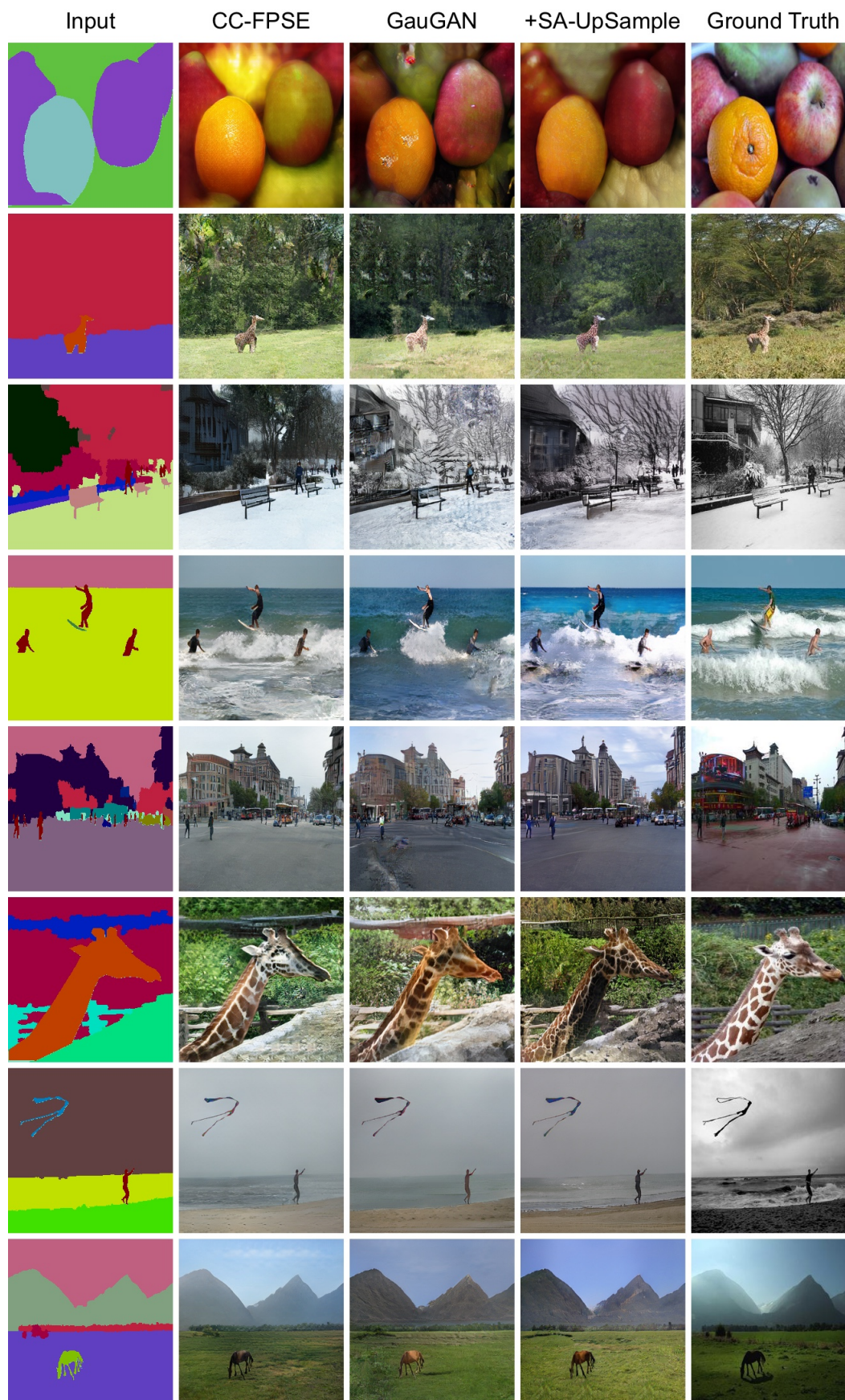


Figure 11: Qualitative comparison on COCO-Stuff. From left to right: Input, CC-FPSE (Liu et al., 2019), GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

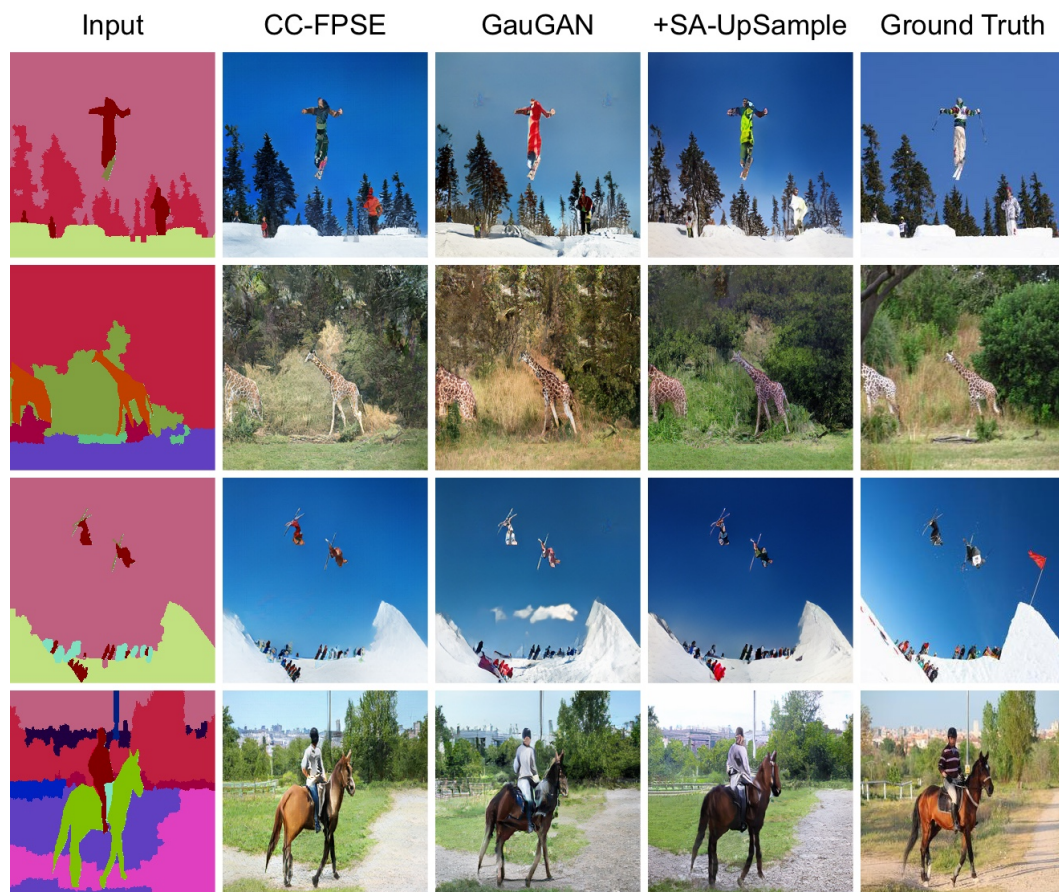


Figure 12: Qualitative comparison on COCO-Stuff. From left to right: Input, CC-FPSE (Liu et al., 2019), GauGAN (Park et al., 2019), GauGAN+SA-UpSample (Ours), and Ground Truth.

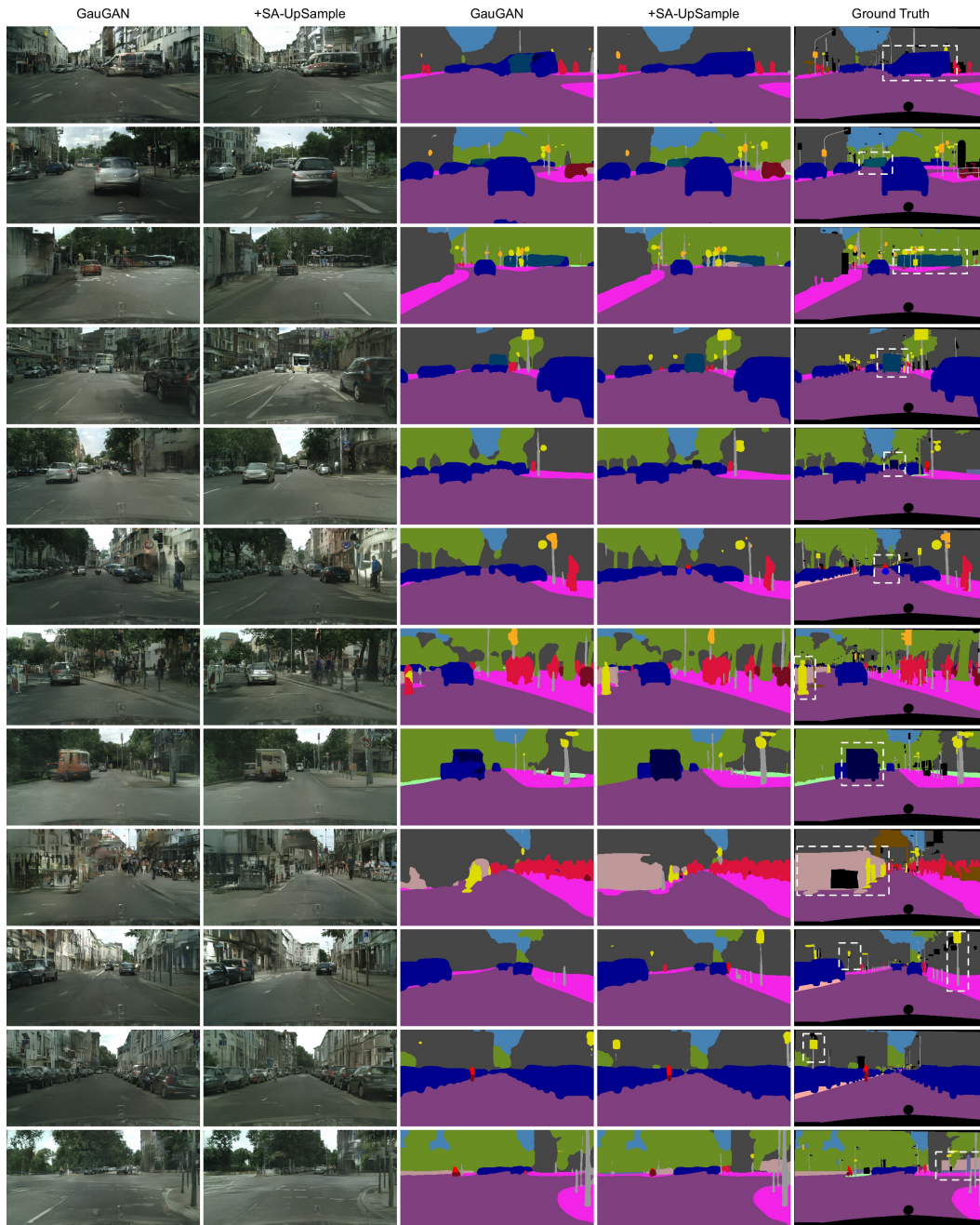


Figure 13: Visualization of our generated semantic maps compared with those from GauGAN (Park et al., 2019) on Cityscapes. Most improved regions are highlighted in the ground truth with white dash boxes.