UGC: Unified GAN Compression for Efficient Image-to-Image Translation

1. Appendix

1.1. Additional Experimental Details

The MACs and parameters of each target student model in our experiment are listed in Table 1 and Table 2.

Table 1. MACs and Params of UGC for different target models on Cityscapes Dataset.

Label Cons.	Model	Arch	MACs	Params	mIoU
10%	Pix2Pix	ResNet	0.866G	0.106M	37.78
			1.398G	0.158M	39.90
		UNet	0.849G	1.173M	44.83
			1.229G	1.071M	44.46
	Pix2PixHD	ResNet	9.735G	4.755M	46.34
			11.595G	7.410M	47.22
	GauGAN	ResNet	8.823G	5.655M	54.85
			12.857G	6.905M	56.50
25%	Pix2Pix	ResNet	0.845G	0.092M	40.90
			1.378G	0.124M	43.53
		UNet	0.734G	1.614M	47.80
			1.235G	2.027M	48.50
	Pix2PixHD	ResNet	9.485G	3.167M	52.40
			13.348G	3.838M	54.42
	GauGAN	ResNet	9.751G	6.283M	60.94
			11.775G	7.928M	61.53
50%	Pix2Pix	ResNet	0.855G	0.091M	42.51
			1.447G	0.142M	44.20
		UNet	0.893G	1.782M	46.10
			1.272G	2.193M	47.23
	Pix2PixHD	ResNet	9.855G	4.231M	56.65
			11.977G	4.410M	57.28
	GauGAN	ResNet	9.507G	5.280M	62.48
			13.354G	3.301M	62.73

1.2. Additional Ablation Study

In this section, we further investigate several important components of UGC. Experiments are conducted on GauGAN[2] and the UNet style generator[1, 3] of Pix2Pix model. The UNet style generator is trained under a MACs constraint of $1.2 \sim 1.3$G on Cityscapes and edges$\rightarrow$shoes datasets.

Table 2. MACs and Params results of UGC in Pix2Pix model on Edges$\rightarrow$Shoes Dataset.

Label Cons.	Arch.	MACs	Params	FID
10%	ResNet	0.866G	0.106M	33.00
		1.398G	0.158M	25.30
	UNet	0.849G	1.173M	37.80
		1.229G	1.071M	34.80
25%	ResNet	0.845G	0.092M	29.04
		1.378G	0.124M	21.43
	UNet	0.734G	1.614M	31.00
		1.235G	2.027M	27.90
50%	ResNet	0.855G	0.091M	24.65
		1.447G	0.142M	19.57
	UNet	0.893G	1.782M	27.15
		1.206G	2.227M	23.25

1.2.1 Analysis of Online Distillation Finetuning.

To demonstrate the effectiveness of our online distillation finetuning procedure in the second stage, we compare the performance of our compressed GauGAN before finetuning. As shown in Figure 1, two observations can be summarized: 1) The first stage of UGC (dubbed as NAS) can obtain a flexible generator that can dynamically change the depth and width while maintaining well performances. Under the same labeled-data constraint, our dynamic GauGAN outperforms the original model with a **16**$\times$ compression ratio. 2) The second stage of UGC (dubbed as NAS+finetune) further boosts the performance of the compressed model, achieving **21.2**$\times$ computational compression and **2**$\times$ labeled-data reduction.

1.2.2 Analysis of Adaptive Data Filter.

To measure the significance of the Adaptive Data Filter (dubbed ADF) in the second stage of UGC, we remove the ADF during the distillation procedure. As is shown in Table 3, Ours w/o ADF gets worse results, which reveals that sub-optimal fake labels generated by the teacher model can mislead the student model and degrade the generation performance. For example, the ADF increases the mIoU from 47.75 to 48.50 on Cityscapes and decreases the FID from 29.96 to 27.90 on edges$\rightarrow$shoes under the 25% labeled data setting. The experimental results demonstrate that our ADF

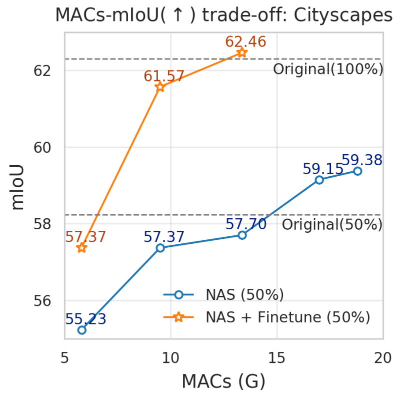

Figure 1. Trade-off curve of GauGAN on Cityscapes. With 50% labeled data, the model trained by the first stage of NAS in UGC outperforms the original model, which shows the potential of UGC in obtaining dynamic models. With the second stage of fine-tuning, UGC achieves comparable performance to the original model with much less computation and label costs.

manages to filter out some noisy images and further improve the performance of the student model.

Table 3. Ablation studies on Adaptive Data Filter.

Dataset	Method	Label Constraints	
		10%	25%
Cityscapes	Ours w/o ADF	43.33	47.75
	Ours	44.46	48.5
Edges→Shoes	Ours w/o ADF	36.09	29.96
	Ours	34.80	27.90

1.2.3 Analysis of Teachers in Distillation.

We illustrate the simplicity and effectiveness of our teacher models choosing strategy during the distillation. Specifically, we use the largest model from the first stage of NAS as the teacher model (dubbed as Largest Teacher), or single teacher search from the search space S (dubbed as Wide/Deep Teacher) to conduct the distillation procedure. Results about different choices of teachers are listed in Table 4. Compared to the largest teacher, our multi-teachers achieve better results. For example, under the 10% labeled setting of Cityscapes, multi-teachers boost the mIoU from 43.74 to 44.46, a reasonable explanation is that the performance gap between teacher and student is too large. Compared to single-teacher distillation, our structural-complementary teachers obtain the best performance and help to break through the bottleneck of capacity for models with low computational costs.

1.3. Additional Qualitative Results

We additionally provide more visualization results of UGC and the state-of-the-art methods in Figures 2, 3 4, 5, 6, 7, which demonstrates the effectiveness of UGC.

Table 4. Ablation studies on Teachers in Distillation.

Dataset	Method	Label Constraints	
		10%	25%
Cityscapes	Largest Teacher	43.74	47.74
	Wide Teacher	43.14	46.43
	Deep Teacher	44.26	46.14
	Multi Teachers	44.46	48.50
Edges→Shoes	Largest Teacher	35.17	28.35
	Wide Teacher	35.22	28.82
	Deep Teacher	36.44	28.41
	Multi Teachers	34.80	27.90

References

[1] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5967–5976, 2017.

[2] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[3] Yuxi Ren, Jie Wu, Xuefeng Xiao, and Jianchao Yang. Online multi-granularity distillation for gan compression. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6793–6803, 2021.

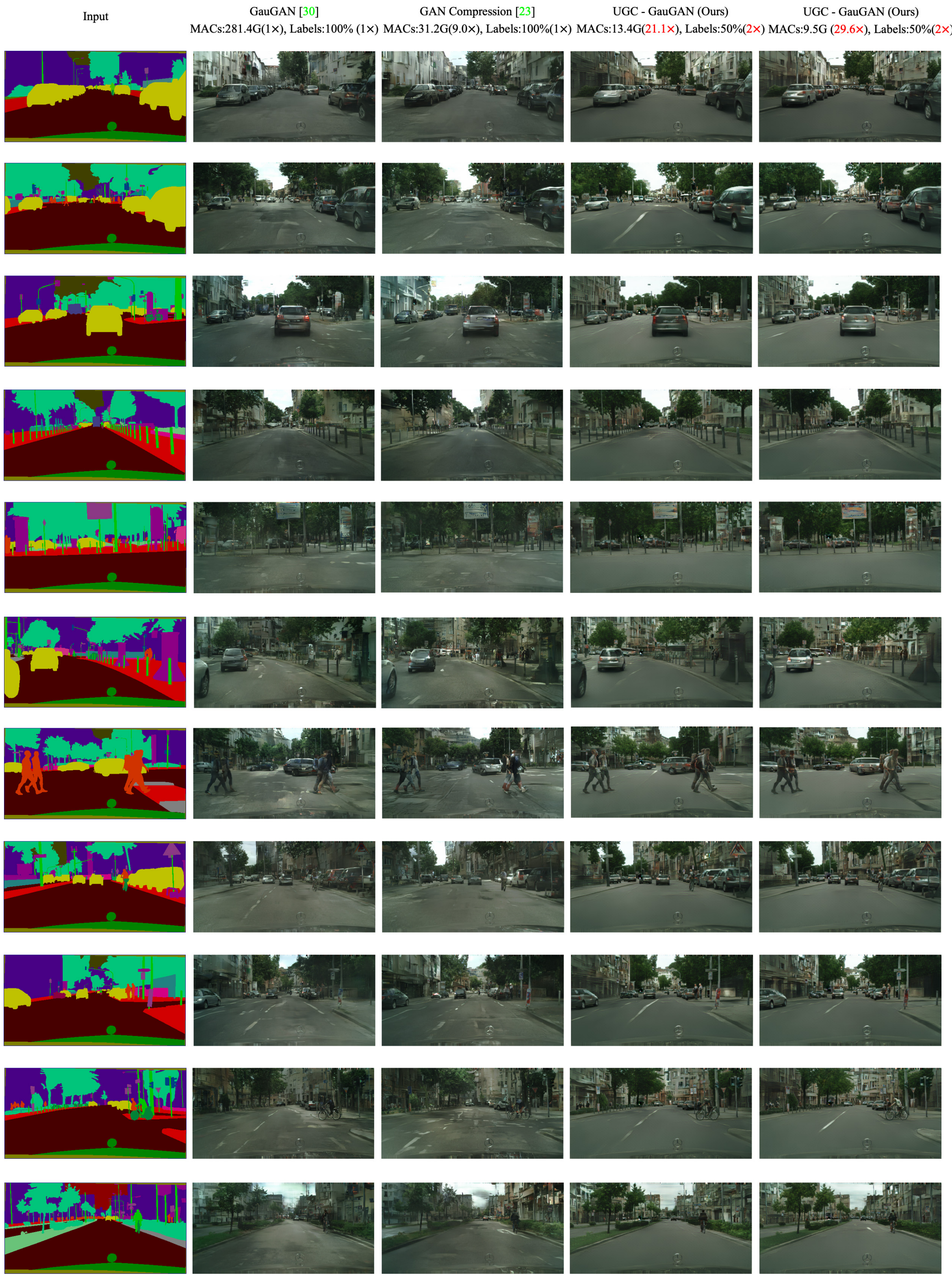

Figure 2. Additional qualitative results of UGC on GauGAN with comparison to the state-of-the-art method and the original model on the Cityscapes dataset.

Input

Pix2PixHD [45]
MACs: 151.3G (1×), Labels: 100% (1×)

UGC – Pix2PixHD (Ours)
MACs: 11.98G (12.6×), Labels : 50%(2×)

UGC – Pix2PixHD (Ours)
MACs: 9.86G (15.4×), Labels : 50%(2×)

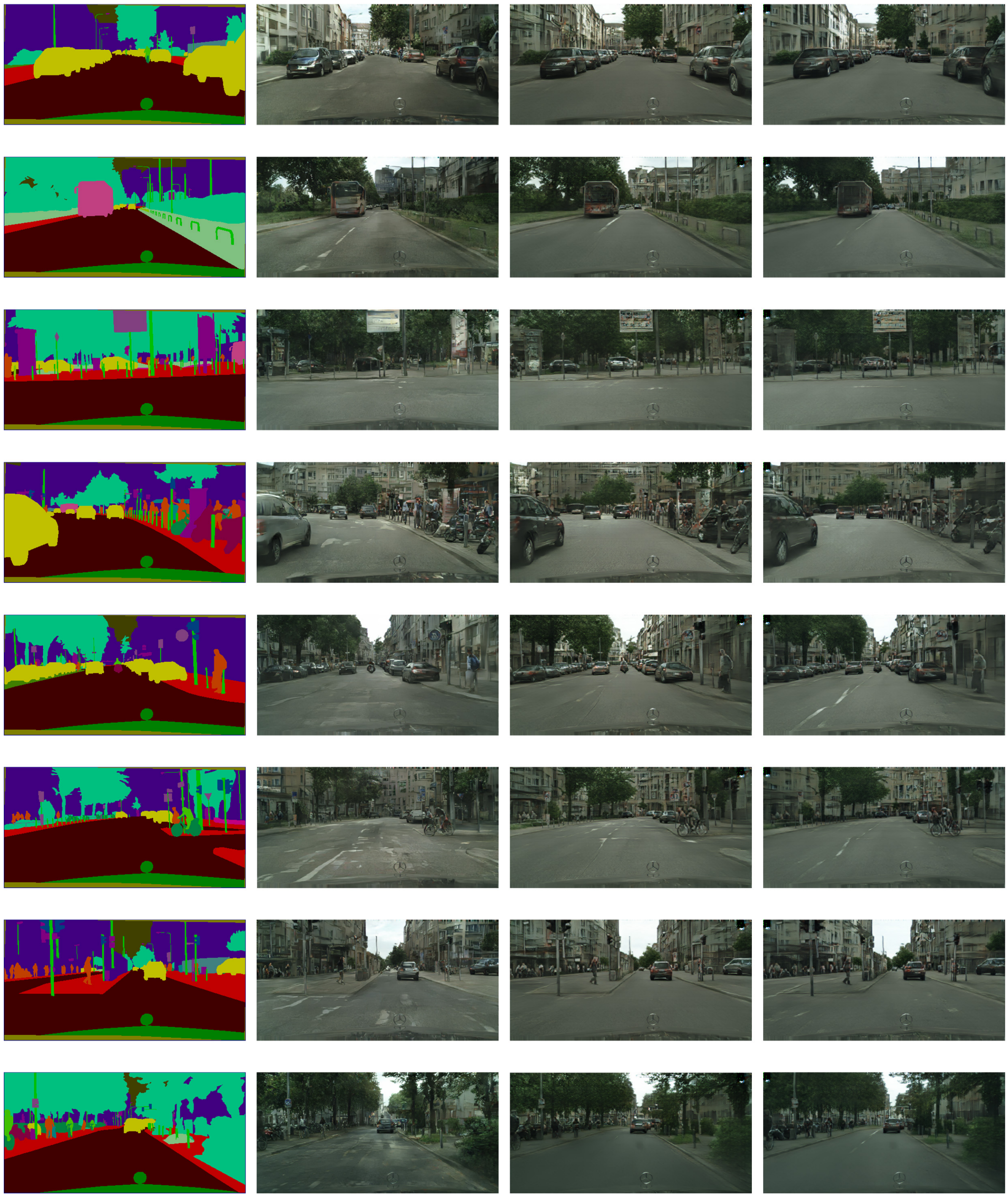

Figure 3. Additional qualitative results of UGC on Pix2PixHD with comparison to the state-of-the-art method and the original model on the Cityscapes dataset.

Figure 4. Additional qualitative results of UGC on Pix2Pix with comparison to the state-of-the-art method and the original model on the Edges∼Shoes dataset. The ResNet style generator is used.

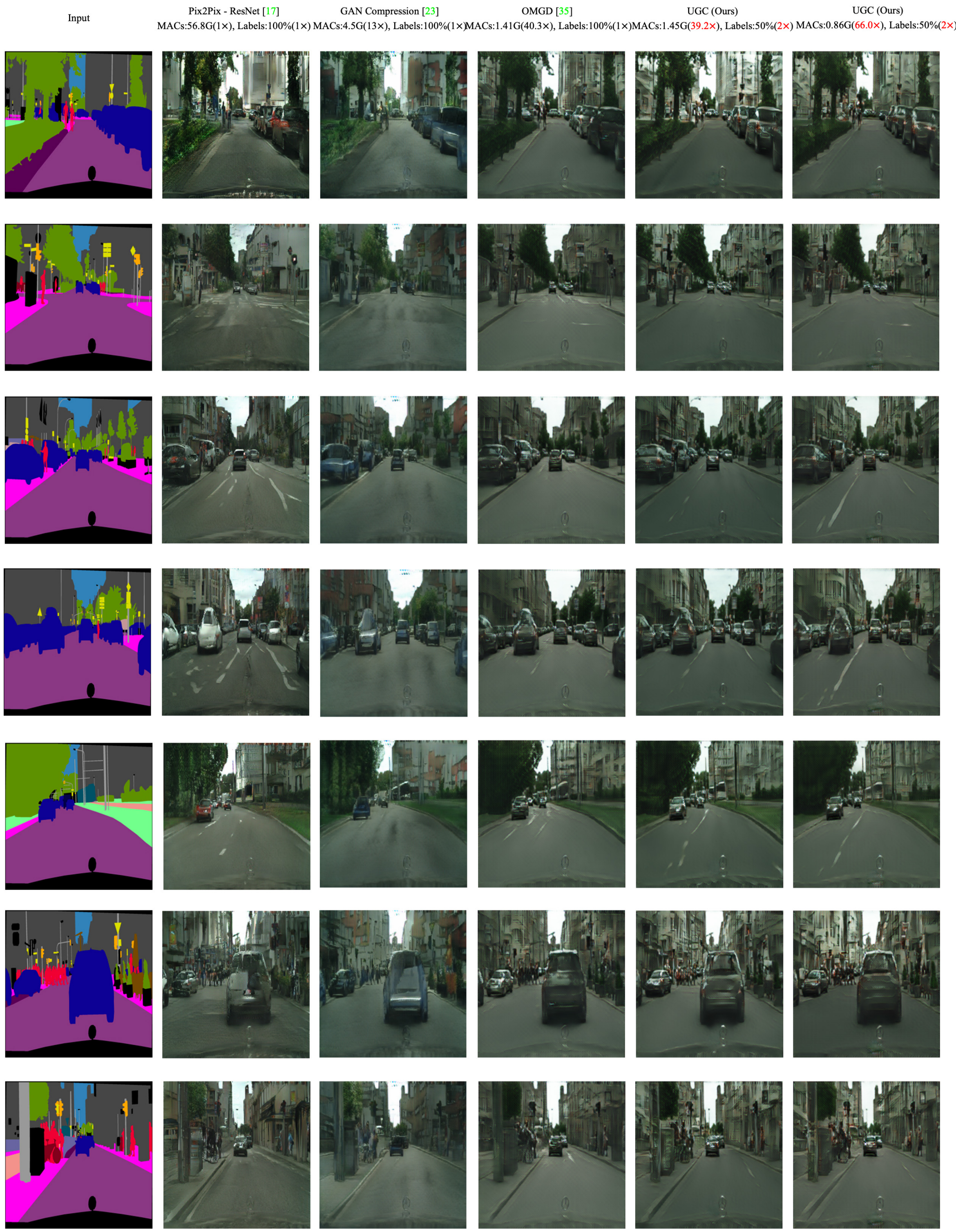

Figure 5. Additional qualitative results of UGC on Pix2Pix with comparison to the state-of-the-art method and the original model on the Cityscapes dataset. The ResNet style generator is used.

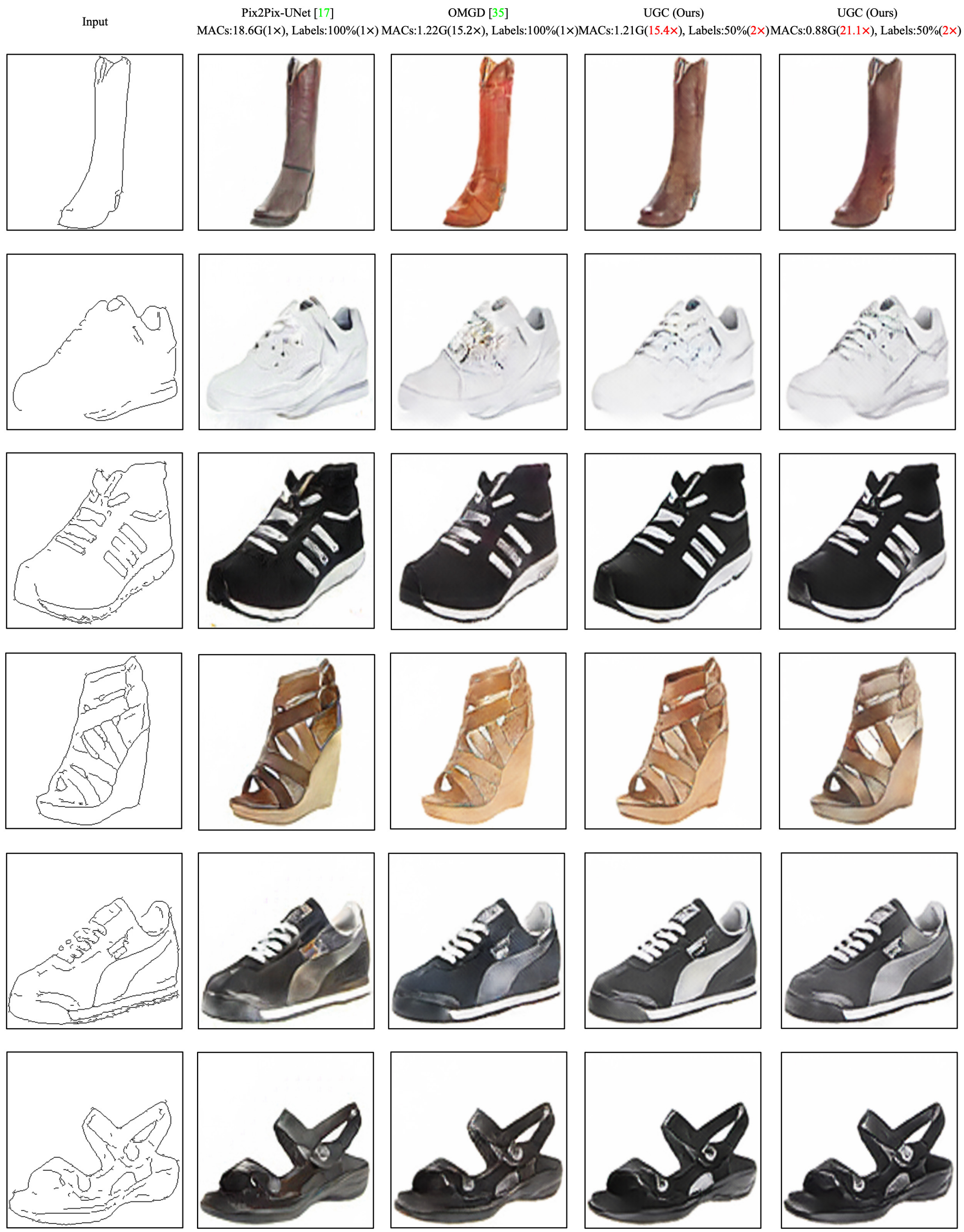

Figure 6. Additional qualitative results of UGC on Pix2Pix with comparison to the state-of-the-art method and the original model on the Edges∼Shoes dataset. The UNet style generator is used.

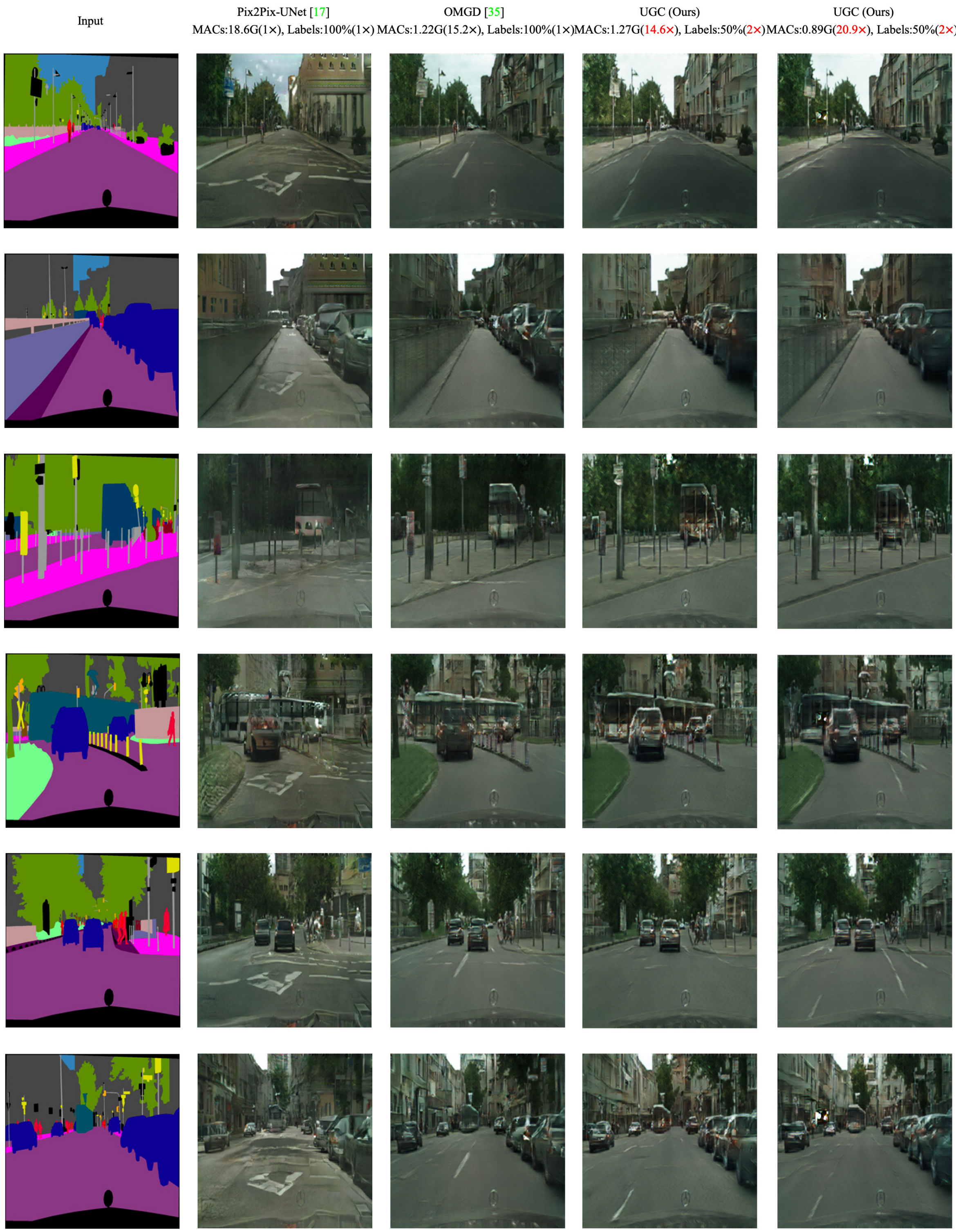

Figure 7. Additional qualitative results of UGC on Pix2Pix with comparison to the state-of-the-art method and the original model on the Cityscapes dataset. The UNet style generator is used.