ScribbleGen: Generative Data Augmentation Improves Scribble-supervised Semantic Segmentation

Supplementary Material

Training Images	FID
10582	43.3
5291	53.7
2646	57.1
1323	58.0

Table 2. FID reported on the validation set. Synthetic images are from our ControlNet model with a varying number of training images. Synthetic images are conditioned on scribbles from the validation set, previously unseen to our model.

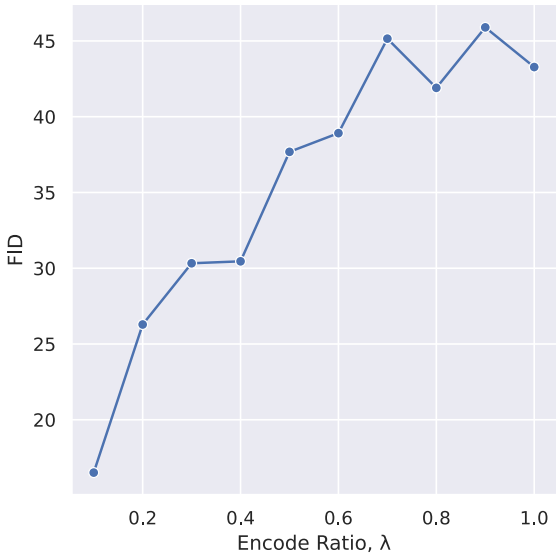

Figure 7. FID reported on the validation set. Synthetic images are from our ControlNet model trained on all of PascalVOC. Images are synthesized conditioned on scribbles from the validation set with varying encode ratios.

6. Validation Set FID Results

In this section, we report the FID of our synthetic images on the validation set. To achieve this, we provide scribbles from the validation set of PascalVOC as conditioning input to our trained ControlNet models. Since the ControlNet models were not trained with data from the validation set, these are previously unseen scribbles. In Table 2, we report the impact of the number of training images of the ControlNet model on validation FID. In Fig. 7, we report the impact of the encode ratio on validation FID.

7. Additional Qualitative Results

In this section, we include additional qualitative results of our method. We provide additional samples of our training data in Fig. 8 and samples from previously unseen scribbles from the validation set in Fig. 9. We also provide visualizations of the effect of the guidance scale on synthesis in Fig. 10 and the effect of the encode ratio in Fig. 11. The effect of the number of training images on synthesis is demonstrated in Fig. 12. Finally, we provide further visualizations of segmentation results in Fig. 13.

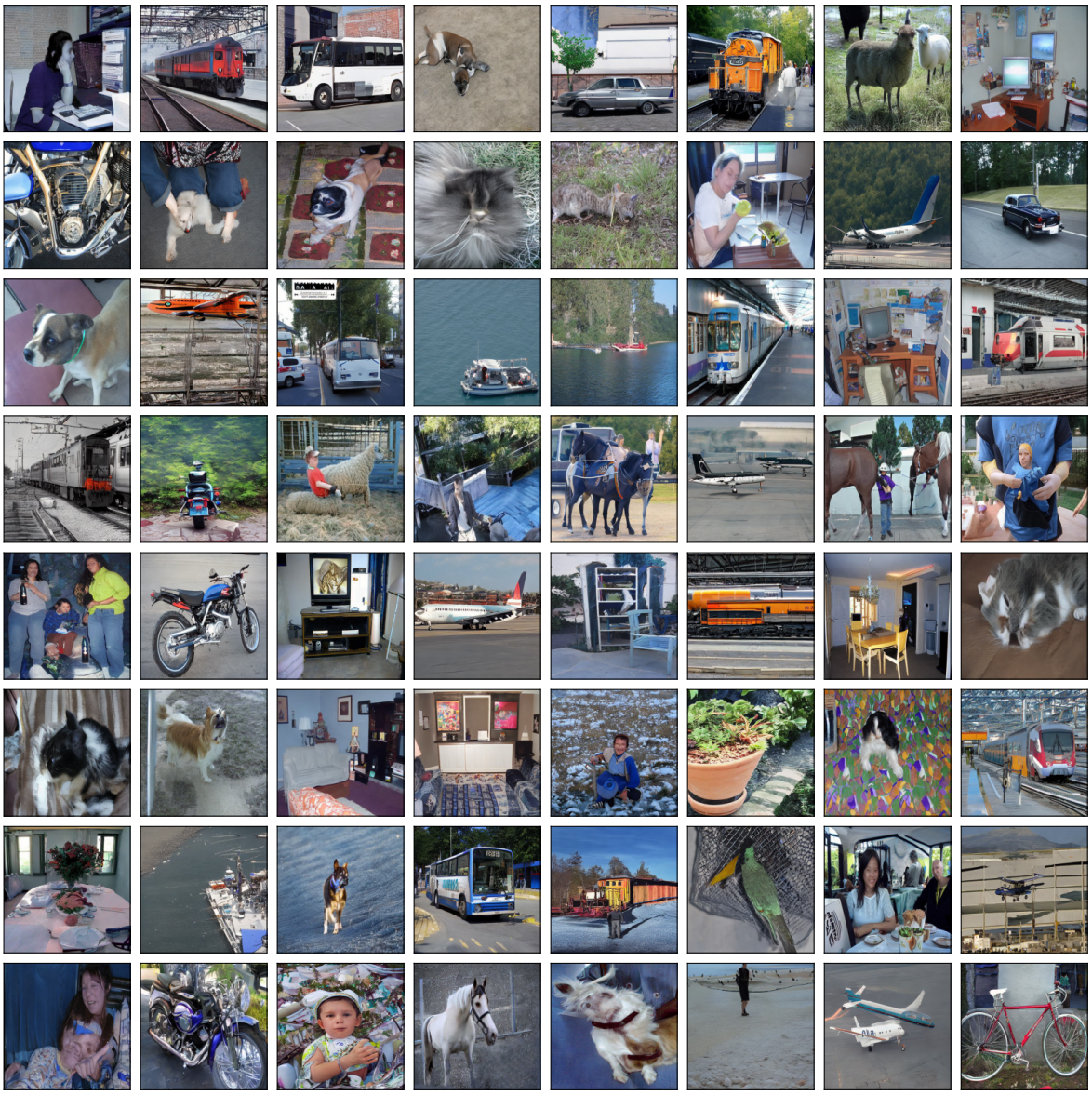

Figure 8. Synthetic training images sampled from a ControlNet model trained on all of scribble-supervised PascalVOC. All images are sampled using guidance scaled $w = 2.0$ and encode ratio $\lambda = 1.0$.

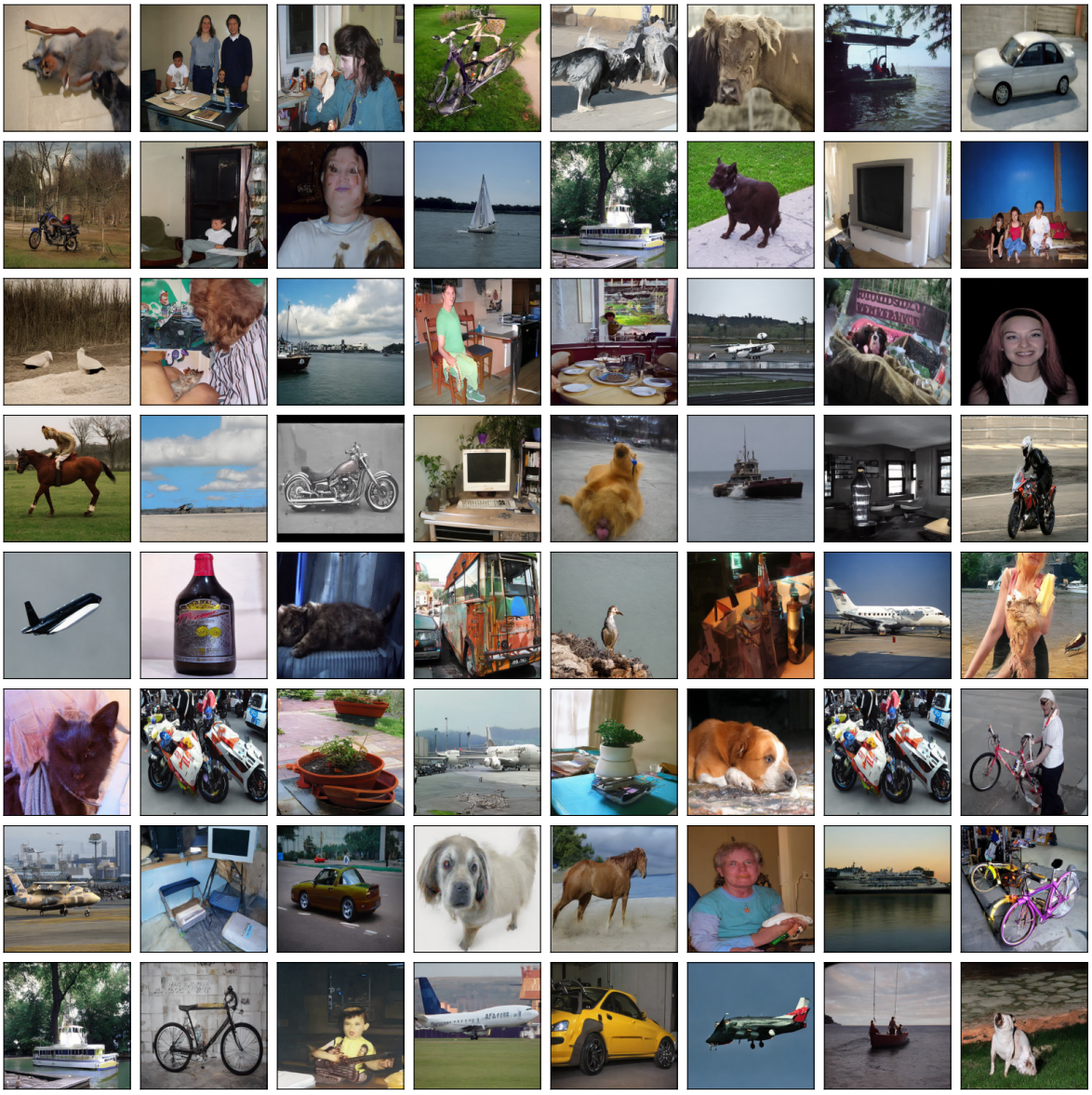

Figure 9. Synthetic validation images sampled from a ControlNet model trained on all of scribble-supervised PascalVOC. Images are synthesized conditioned on scribbles from the validation set, which the ControlNet model has not been trained on. All images are sampled using guidance scaled $w = 2.0$ and encode ratio $\lambda = 1.0$.

Figure 10. Synthetic training images sampled from a ControlNet model trained on all of scribble-supervised PascalVOC. We vary the guidance scale but keep the encode ratio $\lambda = 1.0$ constant to see the effect of the guidance scale on synthesis.

Figure 11. Synthetic training images sampled from a ControlNet model trained on all of scribble-supervised PascalVOC. We vary the encode ratio but keep the guidance scale $w = 2.0$ constant to see the effect of the encode ratio on synthesis.

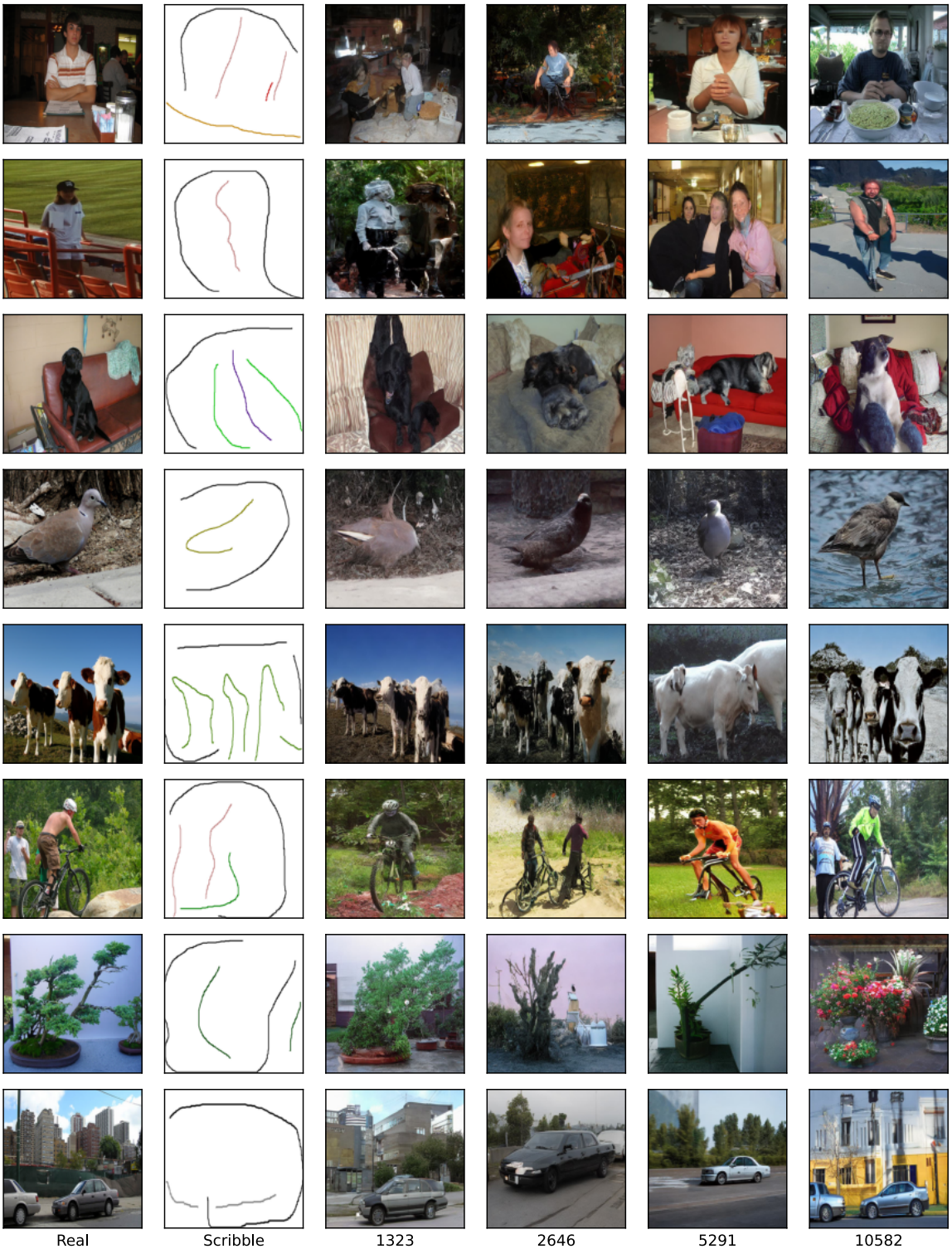

Figure 12. Synthetic training images sampled from a ControlNet model. We vary the number of images on which the ControlNet model is trained to see the impact of the number of training images on synthesis. All images are sampled using guidance scaled $w = 2.0$ and encode ratio $\lambda = 1.0$.

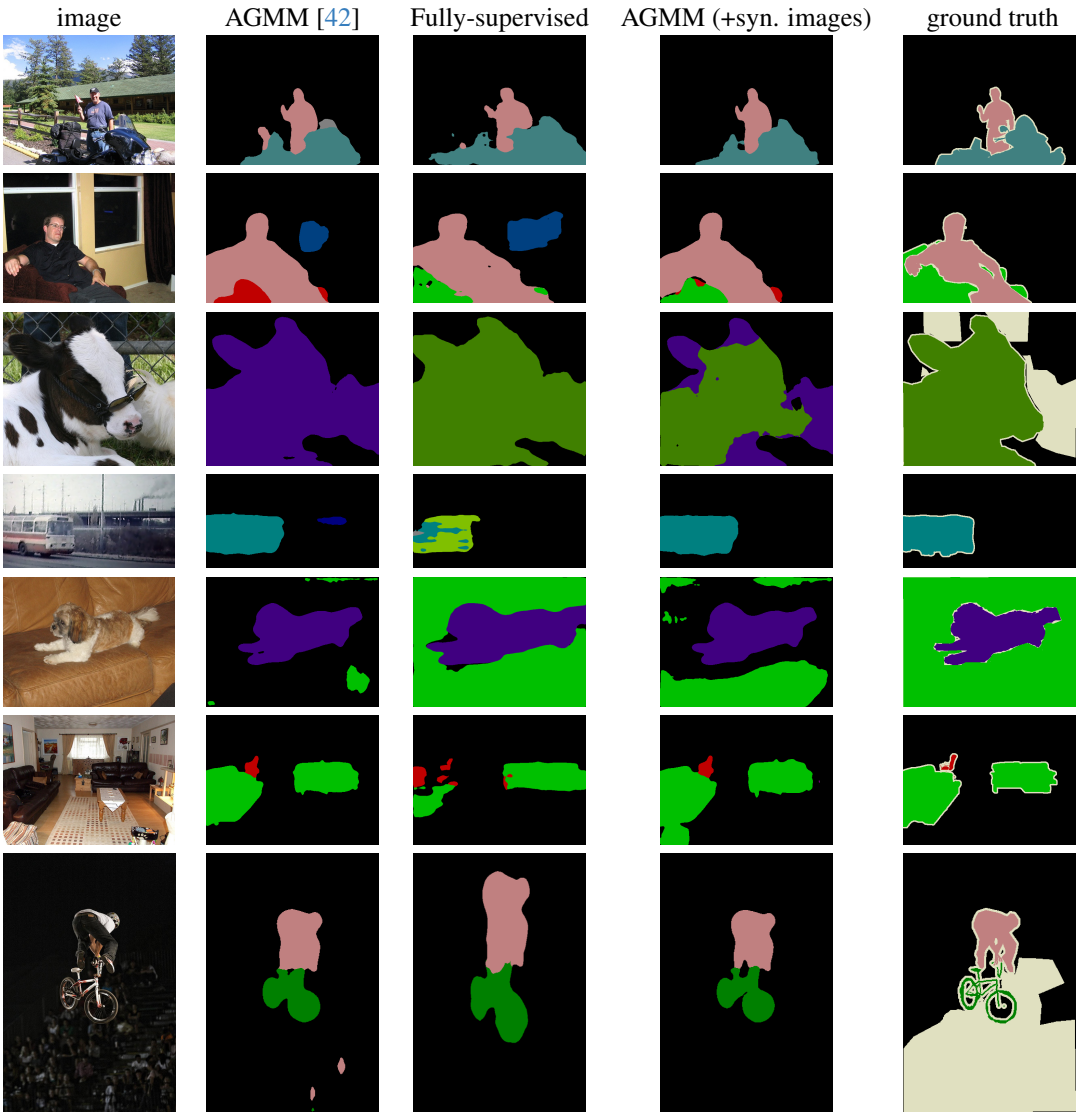

Figure 13. Qualitative results on PASCAL dataset. Our generative data augmentation method improves scribble-supervised semantic segmentation method such as AGMM [42].