

Cascaded Cross MLP-Mixer GANs for Cross-View Image Translation

—Supplementary Material—

Bin Ren¹

bin.ren@unitn.it

Hao Tang²

hao.tang@vision.ee.ethz.ch

Nicu Sebe¹

niculae.sebe@unitn.it

¹ DISI

University of Trento

² Computer Vision Lab

ETH Zurich

This document provides additional experimental results of the proposed Cascaded Cross MLP-Mixer GAN for cross-view image translation task. We present the qualitative results of the ablation studies conducted in our main paper.

1 Qualitative Results of Ablation Study

1.1 Effect of the Number of the CrossMLP Blocks

To figure out how the number of the proposed CrossMLP blocks affects the performance of the final results for our method. Besides the numerical results, we here present the visualization results in Figure 1.

We can see that when the number of the CrossMLP blocks increases to 9, our proposed method achieves the best performance compared to others. We mark out those obvious artifacts for a better comparison. Particularly, the texture and color of grasses (in the third row), the smoothness of the road (in the forth row), and the color of the sky or clouds are more similar to the ground truth images. We conclude that with more CrossMLP blocks used in our experiments, both the geometry structure of a latent mapping pattern and the appearance information relevant to the details objects can be learned progressively. Hence, the final results can be more photorealistic than others.

1.2 Effect of the Refined Pixel-Level Loss

Besides the numerical comparison between B4 and B5 on Dayton-Ablation dataset in our main paper previously. We here provide more visualization results on both Dayton-Ablation and CVUSA dataset for giving further evidence to validate the superiority of our proposed method. In Figure 2, we can see our method with the proposed refined pixel-level loss can generate more realistic house details (the first and the second row). In addition, the detailed things like the texture and style of road, trees, and grasses on Dayton-Ablation dataset becomes more photorealistic when we add the refined pixel-level loss. We can find

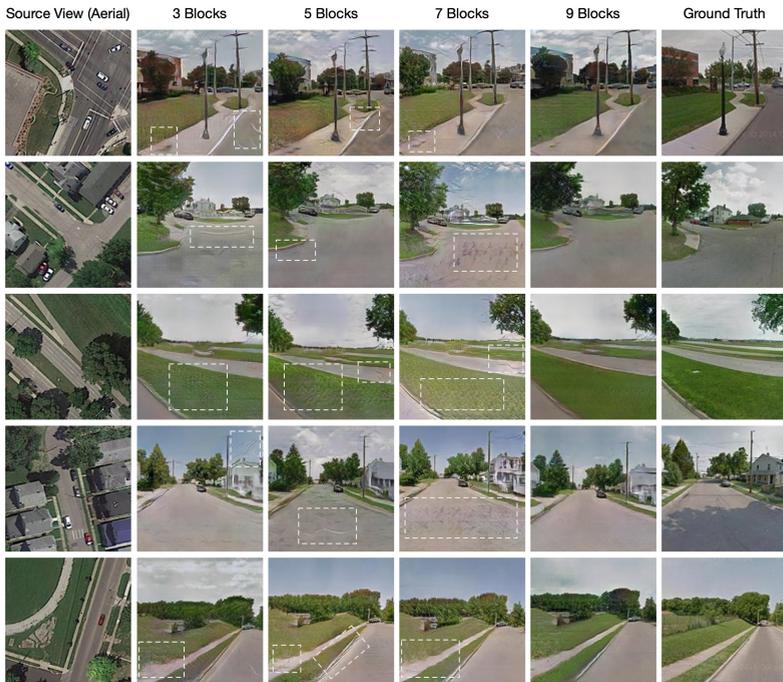


Figure 1: Visualization of the ablation study on the number of the CrossMLP blocks.

the same trend on the CVUSA dataset, too. The color of grasses or sky, the lane line on the road, and the texture of trees or grasses are more photorealistic.

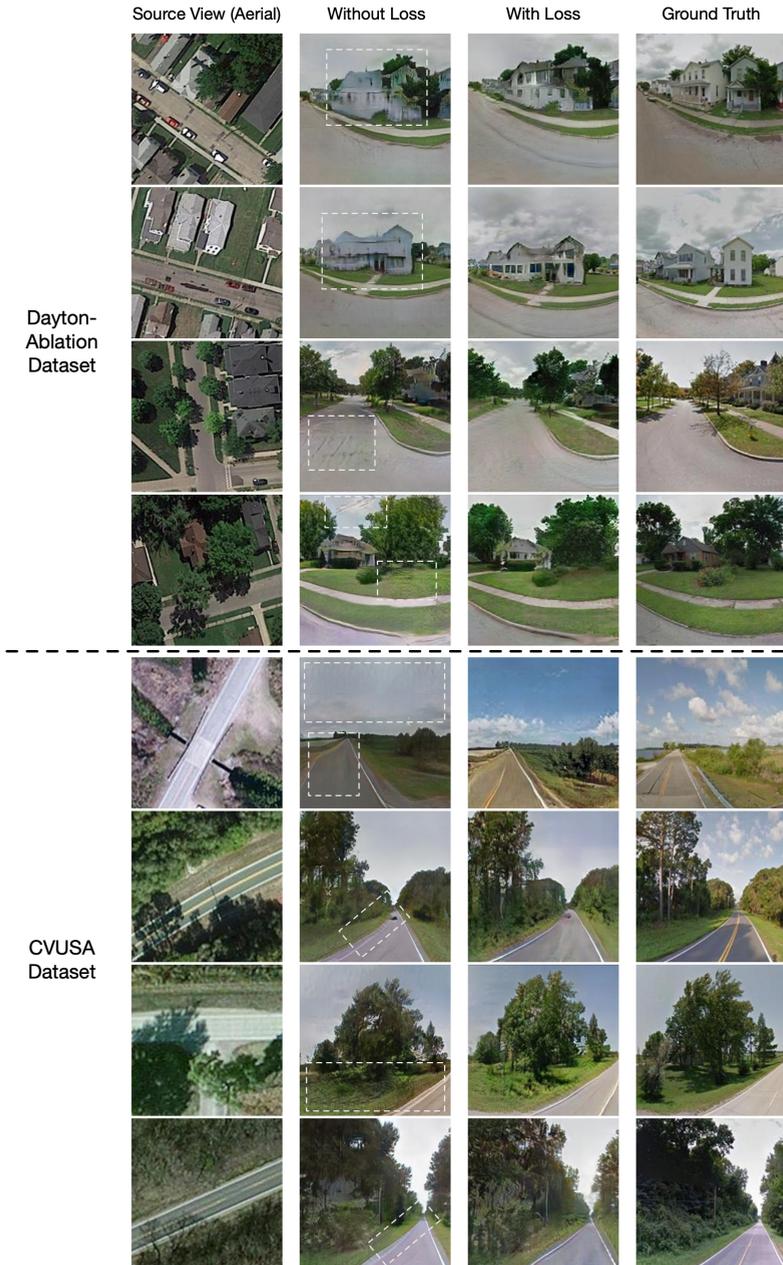


Figure 2: Visualization of the ablation study on the proposed refined pixel-level loss.