Image-to-Image Translation with Conditional Adversarial Networks

Liangjie Cao

July 4, 2018

1. Experiments

To explore the generality of conditional GANs, the authors test the method on a variety of tasks and datasets, including both graphics tasks, like photo generation, and vision tasks, like semantic segmentation:

• $Semantic\ labels \leftrightarrow photo$, trained on the Cityscapes dataset [1].

• $Architectural\ labels \leftrightarrow photo$, trained on the CMP Fa- cades dataset [5].

• $Map \leftrightarrow aerial\ photo$, trained on data scraped from Google Maps.

• $BW \rightarrow color\ photos$, trained on [4].

• $Edges \rightarrow photo$, trained on data from [10] and [7]; binary edges generated using the HED edge detector [6] plus postprocessing.

• $Sketch \rightarrow photo$: tests edges→photo models on human-drawn sketches from [2].

• $Day \rightarrow night$, trained on [3].

Details of training on each of these datasets are provided in the Appendix. In all cases, the input and output are simply 1-3 channel images. Qualitative results are shown in Figures 3, 10, 1, 2, 4, 6, 7, 8, and 5. Several failure cases are highlighted in Figure 9.

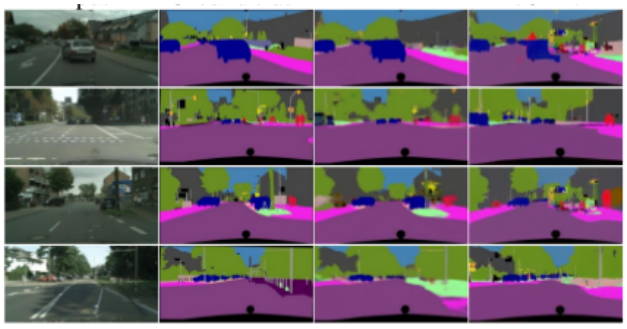

Figure 1. Applying a conditional GAN to semantic segmenta- tion. The cGAN produces sharp images that look at glance like the ground truth, but in fact include many small, hallucinated ob- jects

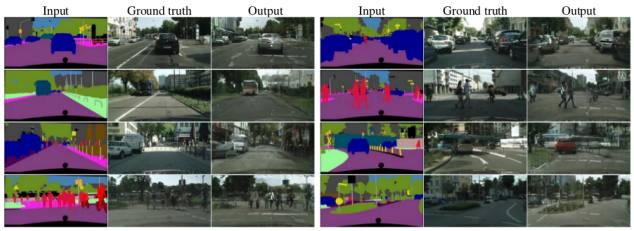

Figure 2. Example results of our method on Cityscapes labels→photo, compared to ground truth

Figure 3. Example results on Google Maps at 512x512 resolution (model was trained on images at 256x256 resolution, and run convolu- tionally on the larger images at test time). Contrast adjusted for clarity

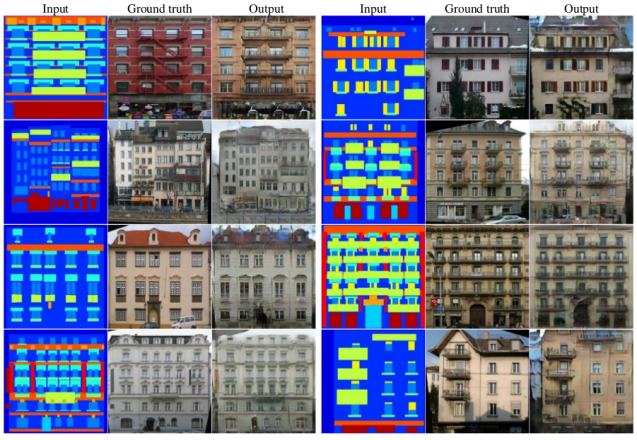

Figure 4. Example results of our method on Cityscapes labelsphoto, compared to ground truth

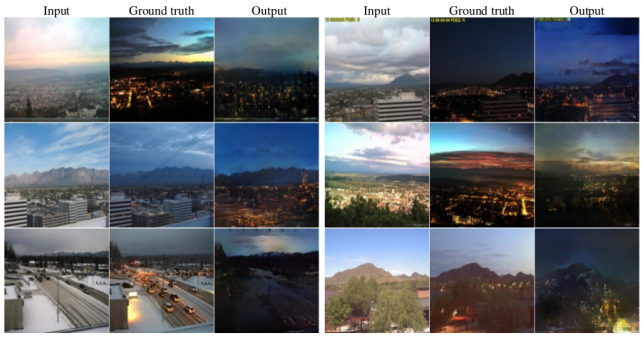

Figure 5. Example results of our method on daynight, compared to ground truth

Figure 6. Example results of our method on automatically detected edgeshandbags, compared to ground truth

Figure 7. Example results of our method on automatically detected edgesshoes, compared to ground truth

Figure 8. Example results of the edgesphoto models applied to human-drawn sketches from [2]. Note that the models were trained on automatically detected edges, but generalize to human drawings

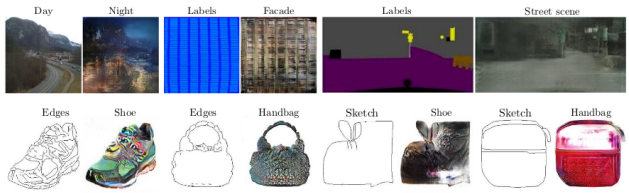

Figure 9. Example failure cases. Each pair of images shows input on the left and output on the right. These examples are selected as some of the worst results on our tasks. Common failures include artifacts in regions where the input image is sparse, and difficulty in handling unusual inputs

Figure 10. Colorization results of conditional GANs versus the L2 regression from [8] and the full method (classfication with rebalancing) from [9]. The cGANs can produce compelling colorizations (first two rows), but have a common failure mode of producing a grayscale or desaturated result(last row)

They note that decent re- sults can often be obtained even on small datasets. Their fa- cade training set consists of just 400 images (see results in Figure 4), and the day to night training set consists of only 91 unique webcams (see results in Figure 5). On datasets of this size, training can be very fast: for example, the re- sults shown in Figure 4 took less than two hours of training on a single Pascal Titan X GPU. At test time, all models run in well under a second on this GPU.

Actually I visited their website to make sure how the model works. Figure 11 and Figure 12 shows what I have done. It's amazing that though my drawing ability is poor, this model can roughly get the key factors of my picture. (To be honest, I draw a rabbit in Figure 12 so the output still

has the features of cats.)

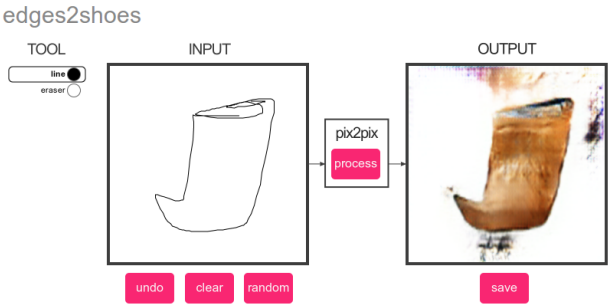

Figure 11. My online testing I

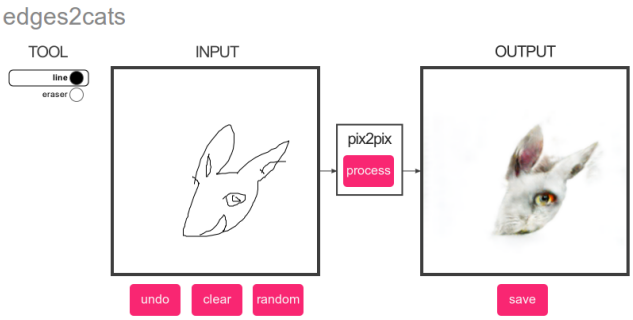

Figure 12. My online testing II

References

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *CVPR*, 2016. 1

[2] M. Eitz, J. Hays, and M. Alexa. How do humans sketch objects? In *SIGGPRAH*, 2012. 1, 2

[3] P. Y. Laffont, Z. Ren, X. Tao, C. Qian, and J. Hays. Transient attributes for high-level understanding and editing of outdoor scenes. *TOG*, 2014. 1

[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, and M. Bernstein. ImageNet large scale visual recognition challenge. *IJCV*, 2015. 1

[5] R. Tyleek and R. ra. Spatial pattern templates for recognition of objects with regular structure. In *GCPR*, 2013. 1

[6] S. Xie and Z. Tu. Holistically-nested edge detection. In *ICCV*, 2015. 1

[7] A. Yu and K. Grauman. Fine-grained visual comparisons with local learning. In *CVPR*, 2014. 1

[8] R. Zhang, P. Isola, and A. A. Efros. Colorful image colorization. In *ECCV*, 2016. 2

[9] Y. Zhou and T. L. Berg. Learning temporal transformations from time-lapse videos. In *ECCV*, 2016. 2

[10] J. Y. Zhu, P. Krhenbhl, E. Shechtman, and A. A. Efros. Generative visual manipulation on the natural image manifold. In *ECCV*, 2016. 1