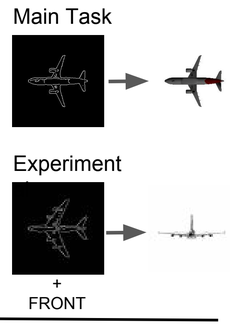
From Sketch to 3D Colored Shading

Yifei Feng, Anna Revinskaya
Stanford University

Introduction

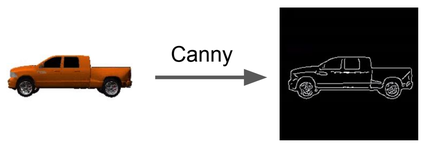
- Generate colored images from edges.
- Investigate the effectiveness of applying various conditional generative adversarial networks (cGAN).
- Produce images of different angles of view for edge images.



Dataset

- ShapeNet [2] : a 3D dataset
- Screenshots of 6 canonical Viewpoints, 4 turntable views

- Cars, ships and planes, total 95170 screenshots
- Use Canny edge detector [1] to generate the sketch



Method

Generative Adversarial Network (GAN) contains two competing neural network models. One (generator) takes noise as input and generates realistic samples. The other model (discriminator) learns to distinguish generated data from real data. Our model architectures are based on conditional GAN (cGAN) [3]. Compared to normal GAN, our generator learns a mapping from noise, conditioned on edge x, to colored image y. Our generator loss is defined as the sum of belows.

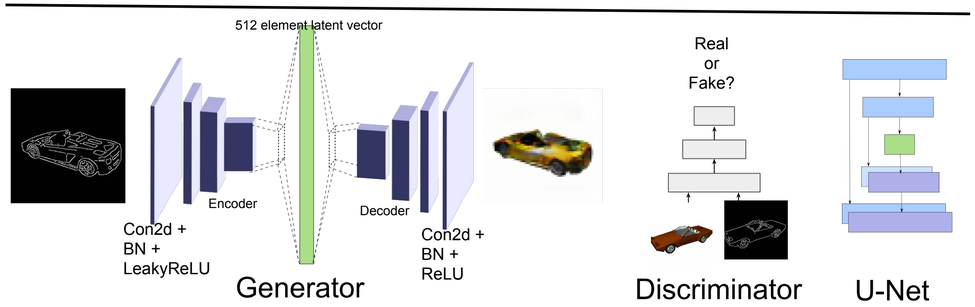
$$L_{G_{orig}} = -E_{x \sim p_{data}(x)}[logD(G(x))]$$

$$L_{L1} = E_{x,y \sim p_{data}(x,y)}[||y - G(x)||_1]$$

Our discriminator loss follows traditional definition:

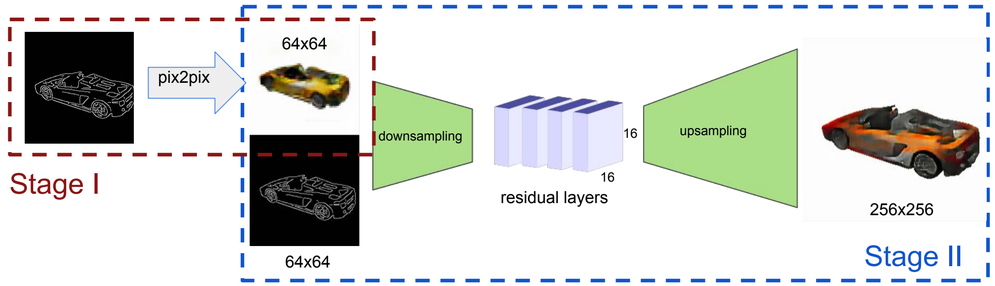
$$L_D = -E_{x,y \sim p_{data}(x,y)}[logD(x,y)] - E_{x \sim p_{data}(x)}[log(1 - D(x, G(x)))]$$

Pix2Pix Model

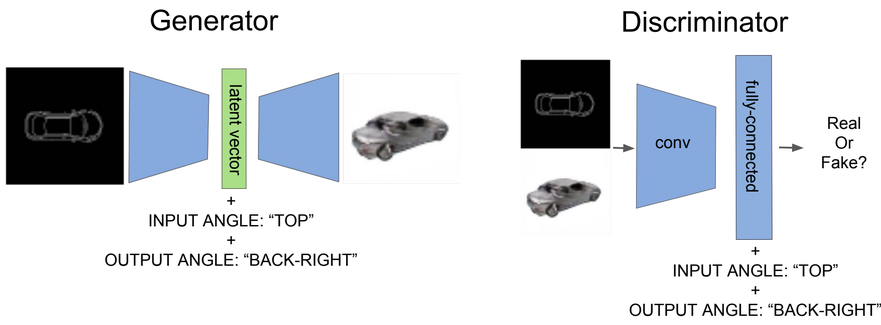


- Dropout is applied both at train and test time to add randomness.
- U-Net architecture adds skip-connections between mirrored encoder and decoder layers.
- Additionally, we used resize convolution for decoder upsampling instead of deconvolution to reduce the checkerboard artifacts [4] in the output.

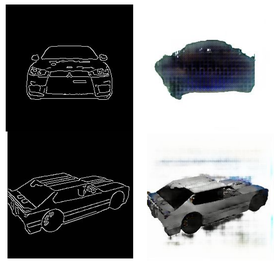
StackGAN Model [5]



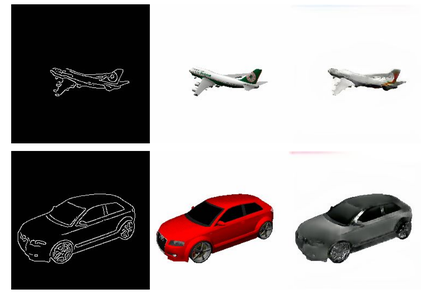
Multi-View Model



Results



Pix2pix model without U-Net(top) does not capture shape of original edge. Deconvolution(down) results with checkerboard artifacts.



Final Pix2pix model with edge(left), groundtruth(middle) and model output(right).

StackGAN results in the order of edge, groundtruth, stage I output and stage II final output from left to right.

Multiview model result with edge(left), groundtruth(middle) and model output(right).

Discussion and Future Work

We implemented two cGAN architectures and experimented with techniques such as U-Net and resize convolutions. We were able to generate colored images from edges with realistic shading and sharp details.

For our bonus task, the Multi-View model often produces blurry images that don’t match the input model. Much less information were provided to the multi-view model due to the lack of U-Net connections and same-view edge input. Increasing the size of our latent vector might help preserve more details. It is also likely that further tuning of hyperparameters would yield better results for the other models as well.

From Sketch to 3D Colored Shading and Multiview Images

Yifei Feng, Anna Revinskaya
Stanford University

References

[1] J. Canny. A computational approach to edge detection. IEEE
Transactions on pattern analysis and machine intelligence, (6):679–698, 1986.
[2] A. X. Chang, T. Funkhouser, L. Guibas, P. Hanrahan, Q. Huang, Z. Li, S. Savarese, M. Savva, S. Song, H. Su, J. Xiao, L. Yi, and F. Yu. Shapenet: An information-rich 3d model repository. Technical Report arXiv:1512.03012 [cs.GR], Stanford University — Princeton University — Toyota Technological Institute at Chicago, 2015.
[3] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. arxiv, 2016.
[4] A. Odena, V. Dumoulin, and C. Olah. Deconvolution and checkerboard artifacts. Distill, 2016.
[5] H. Zhang, T. Xu, H. Li, S. Zhang, X. Huang, X. Wang, and D. Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. arXiv preprint arXiv:1612.03242, 2016.