AUTO3D: Novel view synthesis through unsupervisely learned variational viewpoint and global 3D representation

Supplementary Materials

Paper ID 719

1 Experiment details.

We implemented our model on Pytorch [3]. Our model is trained end-to-end using using ADAM [2] optimization with hyper-parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We used a batch size of 8 for ShapeNet objects. The encoder network is trained using a learning rate of 5×10^5 and the generator is trained using a learning rate 10^4.

2 Discriminator architecture.

Let $\mathcal{C}_{s,k,c}$ denote a convolutional layer with a stride s, kernel size k, and an output channel c. Then, the discriminator architecture can be expressed as $\mathcal{C}_{2,4,32} \rightarrow \mathcal{C}_{2,4,64} \rightarrow \mathcal{C}_{2,4,128} \rightarrow \mathcal{C}_{2,4,256} \rightarrow \mathcal{C}_{1,1,3}$. Note that we use a local discriminator similar to that of [1]. We use a Leaky ReLU activation function with slope of 0.2 on every layer, except for the last layer. Normalization layer is not applied. This architecture is shared across all experiments.

3 Property of Global 3D encoding

Definition 1. Let π be an arbitrary permutation function for a sequence. We say a function $f(\mathrm{X})$ is *permutation-invariant* iff for $\forall\pi$; $f(\mathrm{X}) = f(\pi(\mathrm{X}))$.
Property 1. The Global 3D encoding is *permutation-invariant*, due to the fact that non-local diffusion is permutation-invariant and all features used to calculate $\overline{[f(x)]}$ are considered in an order-less manner (sum or average operation).

4 The Effect of Source Image Ordering

The source images are randomly ordered during training and testing. The goal of this paper is to maximize performance from what is given, not to find the best observation strategy. Although in applications where an agent can actively sample viewpoints, it would be interesting to investigate the effectiveness of observation orderings.

The sum operation used in AUTO3D is essentially permutation invariant. We conduct a simple experiment where we test the model on all possible order. We randomly sampled 1000 tuple of source (image, camera pose) pairs from

ShapeNet cars and chairs, and evaluated on all 24 ordering. We have found that feeding the different order does not affect the performance of proposed AUTO3D. Our model shows robustness to ordering.

5 Additional Results.

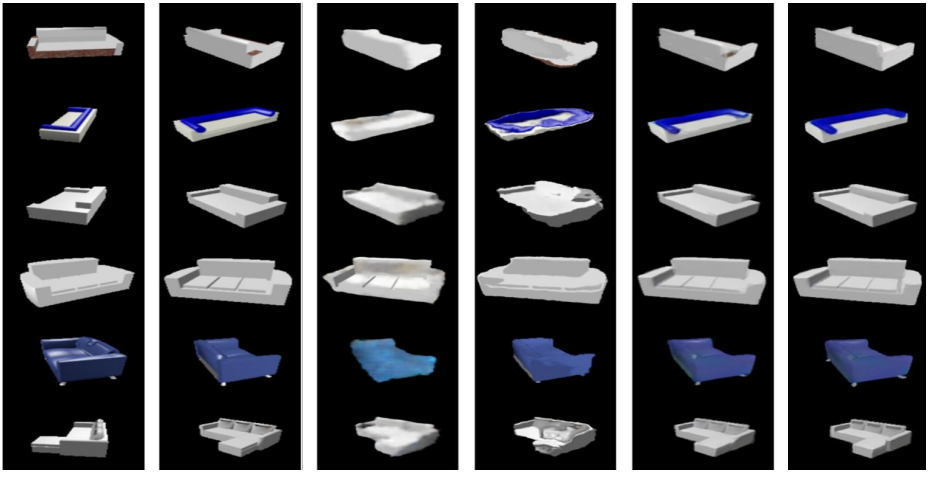

Fig. 1. Comparison of "Sofa" category on ShapeNet with a single 2D input. From left to right: 2D-input, GT, MV3D[4], AF[6], pose-supervised VIGAN[5], Our unsupervised AUTO3D. AUTO3D is comparable to the pose-supervised VIGAN. GT indicates the ground-of-truth.

Code available at: Anonymous ECCV submission.

References

1. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1125–1134 (2017) 1
2. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014) 1
3. Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., Lerer, A.: Automatic differentiation in pytorch (2017) 1
4. Tatarchenko, M., Dosovitskiy, A., Brox, T.: Multi-view 3d models from single images with a convolutional network. In: European Conference on Computer Vision. pp. 322–337. Springer (2016) 2
5. Xu, X., Chen, Y.C., Jia, J.: View independent generative adversarial network for novel view synthesis. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 7791–7800 (2019) 2
6. Zhou, T., Tulsiani, S., Sun, W., Malik, J., Efros, A.A.: View synthesis by appearance flow. In: European conference on computer vision. pp. 286–301. Springer (2016) 2