DeepSCOPE: Unsupervised Geometry Synthesis for Common Objects Pose Estimation

Saksham Gupta, Xiaolong Li, Lynn Abbott

Introduction

Most of the existing 6D pose estimation methods rely on plain CNN with feature pyramid for either key points detection or pose regression, which leads to the main bottleneck being the availability of sufficiently large training datasets with annotations. Weakly-supervised methods that reduce the amount of annotation required are therefore valuable. We aim to find 6D pose of common objects using latent representations learned from unsupervised novel view generation network.

Our contributions include:

- A network that learns implicit 3D representations of objects to generate a novel view of the object given its relative transformation;
- Creating a sizeable synthetic dataset using everyday object's CAD models;
- Increasing the robustness of neural network model towards occlusion by using data augmentation and extra loss function.

Results

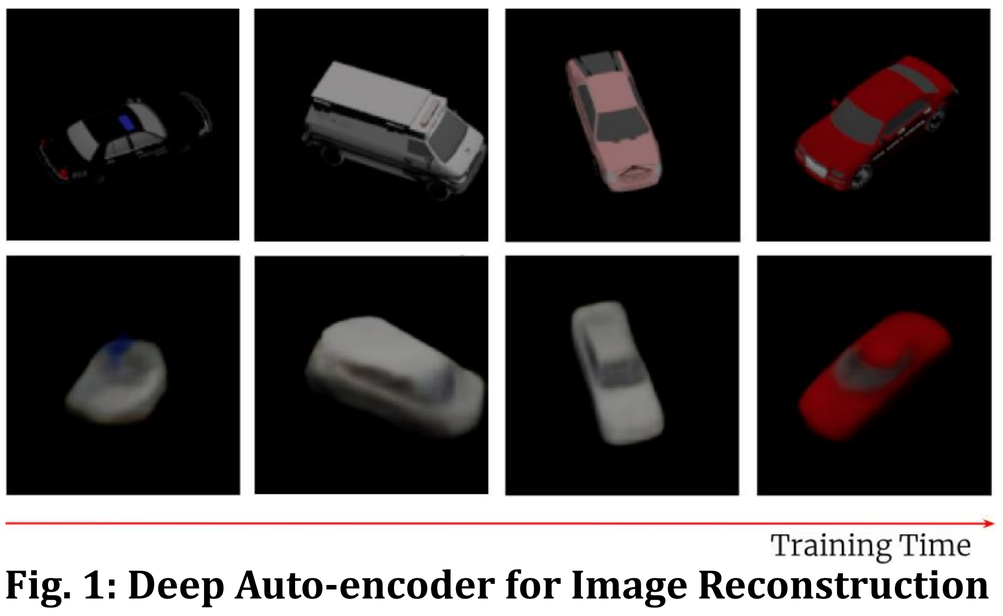

Fig. 1: Deep Auto-encoder for Image Reconstruction

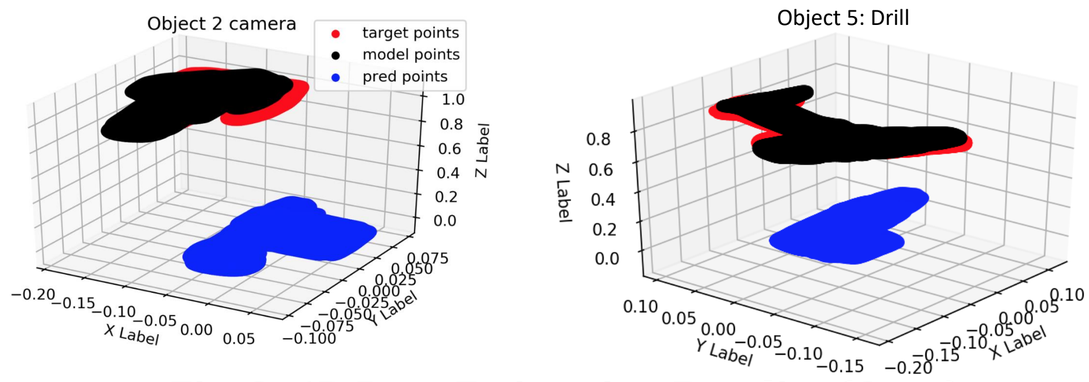

Fig. 2: 6D Pose Estimation Baseline(Ours)

Architecture

Novel View Generation

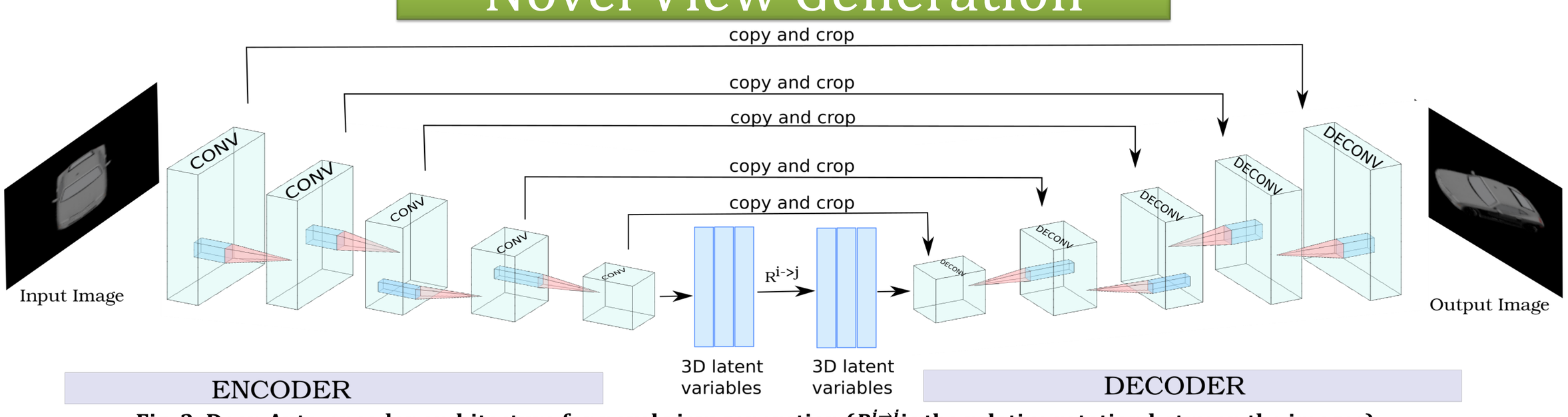

Fig. 3: Deep Auto-encoder architecture for novel view generation ($R^{i \to j}$ is the relative rotation between the images)

Pose Estimation

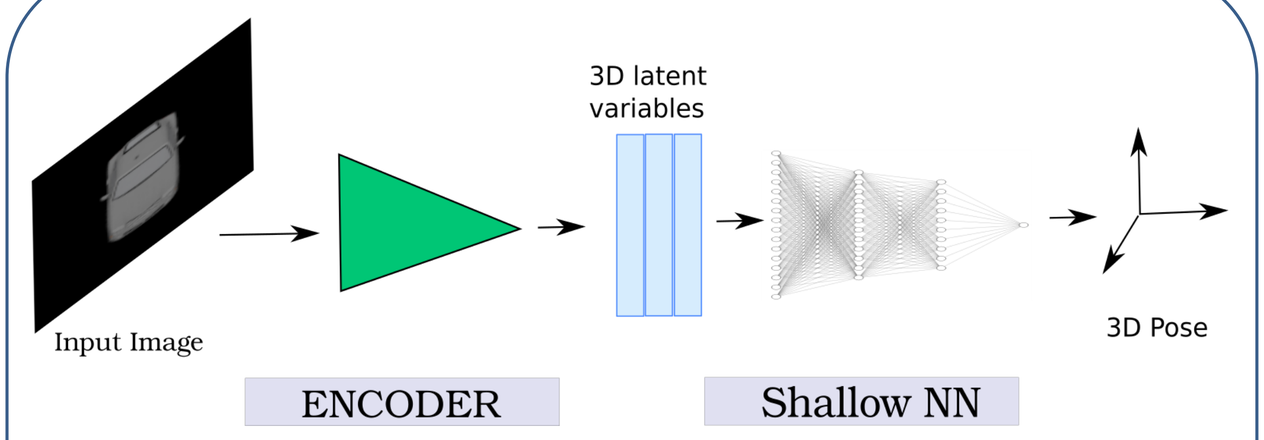

Fig. 4: Pose Estimation architecture. The Encoder is similar to as shown in Fig. 3. To calculate the full 6D Pose of the object a Projective Distance Estimation technique is applied on the detected bounding box of the object

Dataset

Fig. 5: Some examples of *Falling Things* dataset

Conclusion

- Comparable accuracy to the state-of-the-art 6D pose estimation models.
- Pose estimation can be performed using weakly supervised techniques.

Future Work:

- Augment auto-encoders to remove background noise.
- Handle occlusion during novel view generation.

References

[1]. Tremblay, Jonathan, Thang To, and Stan Birchfield. "Falling Things: A Synthetic Dataset for 3D Object Detection and Pose Estimation." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2018.

[2] H. Rhodin, M. Salzmann, and P. Fua, "Unsupervised Geometry-Aware Representation for 3D Human Pose Estimation," *Computer Vision – ECCV 2018 Lecture Notes in Computer Science*, pp. 765–782, 2018.

[3] O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. CoRR, abs/1505.04597, 2015.

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. CoRR,abs/1512.03385, 2015.