Progress Report

Bardia Mojra

February 20, 2021

Robotic Vision Lab

The University of Texas at Arlington

1 To Do

- Implement pose estimation: Implement PVNet, based on ResNet.
- Implement pose estimation: Train PE on YCB data.
- Implement DOPE with added dropout before each layer to estimate variational Bayesian inference (not sure if this is applicable anymore).
- Implement PoseCNN, DOPE, and BayesOD. (No longer applicable - maybe for pose estimation paper?)
- Pose Estimation Servery: Working on different methods, Hollistic, Dense, and Keypoint-based.
- Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers - a review, [1]: Will read after PVNet implementation.

2 Reading List

- [1]
- [2]
- [3]

3 Progress

The following items are listed in the order of priority:

- Pose Estimation: Thank you for this paper. I read and fully annotated the paper. For a single shot, this is exactly what I was thinking about. I have been trying to implement the clean-pvnet version from authors' repository with no avail so. Last night, I installed CUDA-9.0 to run the project but it failed when I tested the provided sample code. When I started my computer this morning, it asked for updates related to GCC-6 and NVidia which are both related to CUDA-9.0. I will try to get to work one more time, then I will move on to the provided Docker implementation. That should be fairly simple. Then, I will start implementing the code as a base, I already have some ideas for

improvements. Their pixel-wise voting scheme has no analytical supporting evidence as to why it should work or why it works. Besides that, I think it is very well written paper. After reading it, I think I have a clear picture of what to write on for the pose estimation survey paper. I started based on PVNet paper, but instead of explanding the related work here, I am developing it under pose estimation survey paper.

- PVNet [4]:Although 6D pose estimation has been the subject of research for many years and great accuracy has been achieved, many states of the art solutions do not take advantage of uncertainty among observed features. In this paper, the authors propose a novel two-stage pose estimation framework, Pixel-wise Voting Network or PVNet, which generates and uses keypoint features uncertainty data. First, they estimate the 2D keypoints for each object in a RANSAC-like fashion which enables uncertainty measurement in the following stage. In the second stage, they use a modified EPnP algorithm [5] that leverages feature uncertainty [3] to calculate estimated object 6D position.

- YCB Dataset [6]: PVNet used a trained ResNet as a base for their model. I will follow their example as they used YCB dataset as well. More importantly, instead of Object Detections, they used ResNet's segmentation feature which allows for pixel-wise voting (and other uncertainty-type implementations).

- Normalized Objects [7]:

- Implement features from PoseCNN, DOPE, and BayesOD.

4 Plans

The following items are listed in the order of priority:

- Pose Estimation in Simulation [8]: Use Nvidia Isaac SDK for in-simulation pose estimation training.

- Look into domain randomization and adaptation techniques.

- Project Alpe with Nolan: On pause for right now.

- UR5e: Finish ROS Industrial tutorials.

5 2021 Goals and Target Journals/Conferences

- Submit a paper on pose estimation with uncertainty to ICIRS.
- Get comfortable with TensorFlow and related Python modules.
- Keep writing.

References

[1] G. Du, K. Wang, S. Lian, and K. Zhao, "Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review," *Artificial Intelligence Review*, pp. 1–58, 2020.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition. corr abs/1512.03385 (2015)," 2015.

[3] L. Ferraz Colomina, X. Binefa, and F. Moreno-Noguer, "Leveraging feature uncertainty in the pnp problem," in *Proceedings of the BMVC 2014 British Machine Vision Conference*, pp. 1–13, 2014.

[4] S. Peng, Y. Liu, Q. Huang, X. Zhou, and H. Bao, "Pvnet: Pixel-wise voting network for 6dof pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4561–4570, 2019.

[5] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision*, vol. 81, no. 2, p. 155, 2009.

[6] B. Calli, A. Singh, A. Walsman, S. Srinivasa, P. Abbeel, and A. M. Dollar, "The ycb object and model set: Towards common benchmarks for manipulation research," in *2015 international conference on advanced robotics (ICAR)*, pp. 510–517, IEEE, 2015.

[7] H. Wang, S. Sridhar, J. Huang, J. Valentin, S. Song, and L. J. Guibas, "Normalized object coordinate space for category-level 6d object pose and size estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[8] Nvidia, "Nvidia isaac sdk — nvidia developer." `https://developer.nvidia.com/Isaac-sdk`, 2021. (Accessed on 02/05/2021).