Progress Report

Bardia Mojra

April 2, 2021

Robotic Vision Lab

The University of Texas at Arlington

1 To Do

- PVNet implementation: Test and document, learn and rewrite.
- Implement pose estimation: Keypoint uncertainty, understand RANSAC.
- Look into methods of generating uncertainty data.
- Pose Estimation Servery: On pause.
- 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

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

3 Progress

The following items are listed in the order of priority:

- Pose Estimation, PVNet [4]: Today, I sat down with Joe, we ran PVNet Docker and tried to run the tests and recreated the same issue. It turns out Cuda 9 is not supported on my GPU (RTX-2060) and he suggested I run it on lab computer. I looked into that, it seems to have weird issue with Cuda again where it thinks it is installed but it cannot find it. Quan suggested I implement the code from scratch and I agree. After recent assignments from Dr. Huber's class I feel much more confident implementing from scratch.

  PVNet architecture is based on a pretrained ResNet-18 [3] and consists of 9 sections, where each section consists of two to three layers. First section consists of a Conv-BN-ReLU followed by a max pooling layer, with skip connection to ninth section. Second section consists of two residual blocks with skip connection to the eighth section. Third section consists of a residual block with stridden convolutions with skip

connection to the seventh block. Fourth section consists of a residual block with stridden convolutions followed by a residual block with a skip connection to the sixth section. Fifth section consists of two residual blocks with dilated convolutions. Sixth section consists of two conv-BN-ReLU layers followed by a bilinear upscaling layer. Seventh and eighth sections each consist of a conv-BN-ReLU layer followed by a bilinear upsampling layer. This network outputs both semantic labels and vector-field predictions estimating center of the object.

I read on RANSAC [5], it is very straight forward. There are many variations of it and I found Neural-Guided RANSAC [6] which allows for optimization of arbitrary task loss functions. They claim it results in large improvement on classic visual tasks.

PyTorch: I started doing some tutorials on PyTorch.

- YCB Dataset [7]: Start with YCB data and look into Berk Calli's work.
- Normalized Objects [8]:
- Implement features from PoseCNN, DOPE, and BayesOD. - On pause.

4 Plans

The following items are listed in the order of priority:

- Pose Estimation in Simulation [9]: 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] 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.

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

[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] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[6] E. Brachmann and C. Rother, "Neural-guided ransac: Learning where to sample model hypotheses," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4322–4331, 2019.

[7] 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.

[8] 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.

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