Progress Report

Bardia Mojra

December 27, 2020

Robotic Vision Lab

The University of Texas at Arlington

1 Long-Term Goals

- Agile Manufacturing

2 Mid-Term Goals

- Bayesian Pose Estimation
- ARIAC
- Pose Estimation Survey

3 To Do

- Dissect Kendall's [1]: in progress.
- Write a literature review for [1]: Done.
- Dissect BayesOD [2]: in progress.
- Write a literature review for [2]: Done.
- Implement PoseCNN, DOPE, and BayesOD.
- Look into transfer learning.
- Read and write a review for [3]: Done.
- Look into domain randomization and adaptation techniques.
- MoreFusion [4]: Wrote a literature review. Done.
- Reading list: [5] and [6].
- Read [7].
- Search for recent pose estimation survey papers.
- Generate new data set for PoseCNN and DOPE using UE4.
- Read more papers on deep pose estimation.
- Learn to use UE4.

4 Progress

Following items are listed in order of priority:

- Pose Estimation with BNN's: I am looking into Yarin Gal's work, he is a leading figure in Bayesian deep learning. His PhD thesis is on Uncertainty in Deep Learning, [8], and was advised by Zoubin Ghahramani. I am thinking if would be a good idea to go over his thesis.

- Reading material: This week I read [9] and [1]. I highlighted important sections of [9] and I taking notes

  Moreover, I

  I am currently working on dissecting [1] and write a review for it simultaneously. Next, I will expand the review I have written for BayesOD and add more details as I look into the source code. My goal is to do some research every morning and to end that with a page of written material. I will focus on Bayesian vision as it offers probabilistic predictions by using per pixel Bayesian inference rather than deterministic output predictions based on MLE parameters as currently used by CNN's.

- Bayesian Pose Estimation: Current pose estimation models with performance comparable with state of the art such as [**?**], [10], [4], and [11] all use CNN's.

  Although the use of CNN's remains a powerful tool as a mean to compute rich feature maps by taking advantage of color channels; it has its short comings, mainly that predictions are deterministic even though the model is probabilistic.

  Classical neural networks and CNN's use maximum likelihood to calculate network weights and biases which derive network outputs. Such models are represented by conditional PDF $p(\mathbf{y}|\mathbf{x},\theta)$ trained on data $D = \{\mathbf{x}_i, \mathbf{y}_i\}_{i=1}^N$ and converging to a set of learned parameters $\theta = \{\mathbf{W}, \mathbf{b}\}$ where

  $$p(D|\theta) = \prod_{i=1}^{N} p(\mathbf{y}_i|\mathbf{x}_i, \theta)$$

defines that PDF. For training these models, parameters **W** and **b** are learned through back-propagation using Maximum Likelihood Estimate (MLE) method, as defined by

$$\hat{\theta}_{MLE} = \underset{\theta}{argMax} \sum_{i=1}^{N} ln\ p(\mathbf{y}_i|\mathbf{x}_i, \theta)$$

but in bayesian —-

We start with the two basic rules for Bayesian machine learning, sum rule and product rule.

Sum rule:

$$p(\mathbf{X}) = \sum_{\mathbf{y}_i} p(\mathbf{y}_i, \mathbf{x}_i)$$

Product rule:

$$p(\mathbf{X}_i, \mathbf{Y}_i) = p(x)p(\mathbf{y}_i|\mathbf{x}_i)$$

Bayesian Learning: Learning is done by applying Bayes' rule to the current state of knowledge as new evidence becomes available throughout the training process.

$$p(\theta|D, M) = \frac{p(D|\theta, M)\ p(\theta|M)}{p(D|M)}$$

Where $p(D|\theta, M)$ is the likelihood of parameters θ in model M. Prior belief or probability of parameters for given model M. And posterior belief of θ given model M and training data D is represented by $p(\theta|D, M)$.

Bayesian Prediction:

$$p(x|D, M) = \int p(x|\theta, D, M)p(\theta|D, M)d\theta$$

Bayesian Model Comparison:

$$p(M|D) = \frac{p(D|M)p(M)}{p(D)}$$

On the other hand, Bayesian Neural Networks or BNN's infer posterior probability distribution frame by frame by marginalizing new likelihood over the distribution of parameters, hence converging to steady-state posterior distribution. Moreover, Bayesian deep learning tools presented in [1] and [8] allows for study and analysis of *aleatoric* and *epistemic* uncertainties together in one frame work.

Aleatoric:

Epistemic uncertainty is what a model does not know due to the incompleteness of training data. Epistemic uncertainty decreases (but it will never be equal to zero) as training data is expanded. It is also referred to as **model uncertainty**.

Learned attenuation:

- ARIAC: For now, I will focus on implementing pose estimation and BayesOD implementations.
- Pose Estimation Survey Paper Feedback: On hold, I am working on Bayesian Pose Estimation.

  Joint Probability: Joint probability $p(D, W)$ relates prior belief to posterior belief given new evidence

  $p(D)p(W|D) = p(D, W)[joint\,prob.] = p(W)p(D|W)[conditional\,prob.]$

  Posterior probability for a particular weight values, W, for given training data D.

  Bayesian approach to deep neural network learning allows for quantafing uncertainties related to model parameters as well as uncertainties rooted in model structure. Uncertainty
- TensorFlow [12]: I am still working through chapter 2.
- Project Alpe with Nolan: On pause for right now.
- UR5e:
- Fellowship:
- System Identification Presentation:

5 Plans

Following items are listed in order of priority:

- (On pause) Continue with ROS Industrial tutorials and documentation.
- (On pause) Resume Robotic Perception course as soon as possible.
- (On pause) Read Digital Image Processing by Gonzalez and Woods.

References

[1] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?," in *Advances in neural information processing systems*, pp. 5574–5584, 2017.

[2] A. Harakeh, M. Smart, and S. L. Waslander, "Bayesod: A bayesian approach for uncertainty estimation in deep object detectors," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 87–93, IEEE, 2020.

[3] J. Tobin, R. Fong, A. Ray, J. Schneider, W. Zaremba, and P. Abbeel, "Domain randomization for transferring deep neural networks from simulation to the real world," in *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 23–30, IEEE, 2017.

[4] K. Wada, E. Sucar, S. James, D. Lenton, and A. J. Davison, "Morefusion: Multi-object reasoning for 6d pose estimation from volumetric fusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14540–14549, 2020.

[5] J. Lampinen and A. Vehtari, "Bayesian approach for neural networks—review and case studies," *Neural networks*, vol. 14, no. 3, pp. 257–274, 2001.

[6] R. Li, Z. Liu, and J. Tan, "A survey on 3d hand pose estimation: Cameras, methods, and datasets," *Pattern Recognition*, vol. 93, pp. 251–272, 2019.

[7] C. Choi and H. I. Christensen, "Rgb-d object tracking: A particle filter approach on gpu," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1084–1091, IEEE, 2013.

[8] Y. Gal, "Uncertainty in deep learning," *University of Cambridge*, vol. 1, no. 3, 2016.

[9] N. Sünderhauf, O. Brock, W. Scheirer, R. Hadsell, D. Fox, J. Leitner, B. Upcroft, P. Abbeel, W. Burgard, M. Milford, *et al.*, "The limits and potentials of deep learning for robotics," *The International Journal of Robotics Research*, vol. 37, no. 4-5, pp. 405–420, 2018.

[10] J. Tremblay, T. To, B. Sundaralingam, Y. Xiang, D. Fox, and S. Birchfield, "Deep object pose estimation for semantic robotic grasping of household objects," *arXiv preprint arXiv:1809.10790*, 2018.

[11] C. Wang, D. Xu, Y. Zhu, R. Martín-Martín, C. Lu, L. Fei-Fei, and S. Savarese, "Densefusion: 6d object pose estimation by iterative dense fusion," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3343–3352, 2019.

[12] B. Planche and E. Andres, "Hands-on computer vision with tensorflow 2," 2019.