Progress Report

Bardia Mojra

January 21, 2021

Robotic Vision Lab

The University of Texas at Arlington

1 To Do

- Implement DOPE with added dropout before each layer to estimate variational Bayesian inference.
- Finish Tensorflow tutorials, [1]: 4/9 chapters done.
- Learn to implement transfer learning.
- Finish Docker tutorials, [2]: 4/18 chapters done.

2 Progress

Following items are listed in order of priority:

- [3]: In recent years dropout has become a usefull tool in deep learning as [4] and [5] demonstrated its potentials as well as mathematical derivation explaining its internal mechanism. In this paper, authors expand on their previous work and further define uncertainty in Bayesian deep CNN's. They divide uncertainty to two groups, Aleatoric and Epistemic. Aleatoric uncertainty is the observation noise within the system, this could be camera noise or mobile agent's structural vibration that induces noise on images. Epistemic uncertainty is the model uncertainty and converges to zero given expanded dataset. Furthermore, they categorize Aleatoric uncertainty into two groups, Heteroscedastic and Homoscedastic, which represent input-dependent and input-independent noise, respectively. I believe further system identification can be performed efficiently on robotic vision systems, especially monocular mobile vision systems. As plainly mentioned in the paper, they use Intelligent Robust Control (IRC) ideas with Bayesian convnets to infer a posterior distribution for model f, for some input image x and a predicted output mean and variance $\hat{y}$ and $\hat{\sigma^2}$ (variance is interpreted as aleatoric uncertainty). Per Intelligent Robust Control, output predictions are then fed into a inner feedback loop where model parameters, W, are updated, $\hat{W}$, using latest prediction errors. This is why heteroscedastic uncertainty is referred to as *learned loss attenuation*, because in Intelligent robust control effects of a prior input on the current output is dieluted if it does not become a pattern.

- [6]: This paper lays out some of the most important challenges ahead of researchers in deep learning for robotics, and in particular robotics vision. Authors present these challenges as three conceptually orthagonal axes, *learning*, *embodiment*, and *reasoning. Learning* challenges are described by the need for intelligent agents to learn actively on their own if needed request operator assistance. Authors believe, with recent development in deep learning, uncertainty estimation can be used to detect never seen before objects and assign new labels and classes as needed. This is called Active Learning and it is a prime goal of machine learning researcher. *embodiment* challenges are described by the need for vision systems and robotic agents to better understand and ultimately engage with their environments in an intuitive way. Temporal and spatial embodiments are used to encode higher dimension information about a pixel's state over time and its relation to other pixels. Embodiment of further information extracted from the environment, if done with computational efficiently, can make significant progress towards active vision and manipulation as we could draw posterior probability from them. *Reasoning* challenges are characterized by the ever growing need for intelligent agents to perform more sophisticated reasoning. This is investigated at three tiers, *reasoning about object and scene semantics*, *reasoning about object and scene geometry*, and *joint reasoning about semantics and geometry* and the idea is to take advantage many semantic regularities present in real world by using prior knowledge. In psychology, this is generally referred to as *schema inference* and schema is defined as a framework or a pattern of thought and behavior that organizes information and relation among them. In robotic vision, similar to human behavior, agents need to make shortcut inferences bases on more prominant prior knowledge and current environmental and situational needs. This is also referred to as pixel-to-action learing. Lastly, authors pose the question whether it is more feasible to pursue model based learning or deep learning for future research. It is reasonable to belief systems will continue to develop somewhere in the middle of the spectrum rather than the two ends.

- Object Pose Estimation with Uncertainty: Right now I am trying to put together BayesianOD and PoseCNN.

- Implement PoseCNN, DOPE, and BayesOD.

- Transfer learning is a powerful tool in deep learning for reusing pre-trained models (weights and parameters) for training other models. This is done by removing high level layers and replace them with those of the new model. Then we freeze the weights of pre-trained model and train the new layers on new target data.

- Look into domain randomization and adaptation techniques.
- Search for recent pose estimation survey papers: Found this paper, [7], on pose estimation.
- Bayesian Pose Estimation Notes: Current pose estimation models with performance comparable with state of the art such as [8], [9], [10], 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 $\mathbf{W}$ and $\mathbf{b}$ 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 [3] and [12] 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:

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

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

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

Bayesian approach to deep neural network learning allows for quantifying uncertainties related to model parameters as well as uncertainties rooted in model structure. Uncertainty

3 Plans

Following items are listed in order of priority:

- Implement multiple object pose estimation with uncertainty estimation.
- Keep working on Bayesian Pose Estimation paper.
- ARIAC: For now, I will focus on implementing pose estimation and BayesOD implementations.
- Continue on UE4 tutorials.
- Pose Estimation Survey Paper Feedback: On hold, I am working on Bayesian Pose Estimation.
- Project Alpe with Nolan: On pause for right now.
- UR5e: Finish ROS Industrial tutorials.

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

[2] G. Schenker, *Learn Docker – Fundamentals of Docker 19.x: Build, test, ship, and run containers with Docker and Kubernetes, 2nd Edition*. Packt Publishing, 2020.

[3] 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.

[4] Y. Gal and Z. Ghahramani, "Dropout as a bayesian approximation: Representing model uncertainty in deep learning," 2016.

[5] A. Kendall and R. Cipolla, "Modelling uncertainty in deep learning for camera relocalization," 2016.

[6] 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.

[7] 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.

[8] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox, "Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes," 2018.

[9] 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.

[10] 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.

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