# Towards Robust Autonomous Semantic Perception

Yuri Feldman and Vadim Indelman

Object detection and classification is a component of situational awareness basic to numerous robotics applications involving semantic world perception and mapping. While object classification is considered largely solved in many ”controlled” computer vision scenarios, challenges often remain in applications which involve mobile robots, especially where semantic information affects autonomous decision making. The latter often require algorithms to function under rather general assumptions on robot environment and its localization within it, maintaining and refining awareness of the environment while imaging complex scenes under uncertain robot motion. Challenges include partial or full object occlusions, class aliasing (due to classifier imperfections or objects that appear similar from certain viewpoints), imaging problems, false detections.

The mobility of robotic systems is widely exploited to overcome some of these challenges by accumulating classification evidence across multiple observations and viewpoints [1], [2], [8], [10], [12], [13], [17], [19], including a recent surge in active methods for autonomous classification, where next viewpoints are automatically selected, e.g. [1], [2], [8], [13], [17], [19]. Variations in object appearance are often directly addressed using offline-built class models for inference rather than raw classifier measurements. Especially in the active methods, such models are often themselves spatial and view-dependent. As was shown by Teacy et al. [17] and Velez et al. [19] view-dependent models can allow for better fusion of classifier measurements by modelling correlations among similar viewpoints instead of the common but usually false assumption of independence of measurements.

Reliance on spatial models however introduces new problems, as robot localization is usually not precisely resolved, leading to errors when matching measurements against the model. This is aggravated in the presence of classifier measurements actually not complying to the model, as may happen for example when a classifier is deployed in an environment different in appearance from the one it was trained on, for example - in another country where objects semantically identical to the ones in the training set look differently. In the latter case, classifier output would often be arbitrary, rather than reflect the actual uncertainty in classification, known as *epistemic* or *model uncertainty* [6], [9]. In the domain of Bayesian deep learning, methods exist to approximate the above as network posterior [3], [6], [11], for example using test-time dropout [5], which allows to (approximately) obtain it for virtually any deep learning-based classifier without change in model (Fig. 2).

Accounting for uncertainty is directly related to novelty detection and safety e.g. [7], [15], and confidence prediction [16], [18]. It essentially allows the system to be aware of low confidence situations and avoid autonomously making confident wrong decisions, in the case of classification - assigning a wrong class with high confidence.

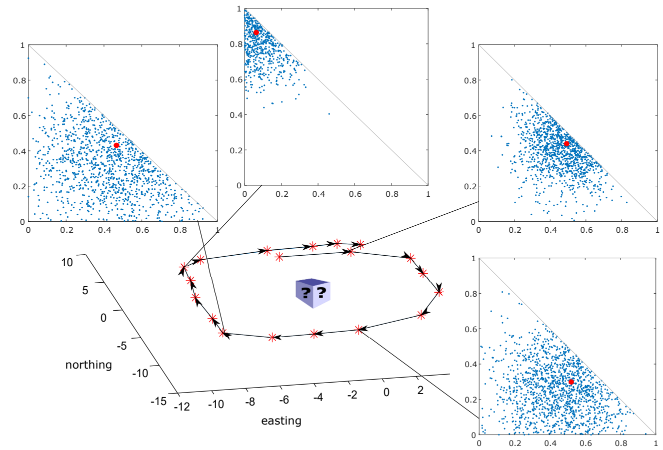



**Fig. 1:** Robot acquires observations along track in the vicinity of the object of interest. At each time step, classifier outputs a cloud of classification vectors reflecting the model uncertainty, unlike a single vector measurement (red dot) or a component thereof in many current approaches.

Existing classification fusion methods however do not address model uncertainty. Indeed, with few exceptions most current methods discard also the classification vector commonly output by the classifier, only using the most likely class (component with highest response) for belief update. Likewise, most methods ignore uncertainty in localization, assuming it perfectly known.

In light of the above, we develop a method [4] for fusing responses of a classifier which provides a model uncertainty measure, while accounting for viewpoint-dependent variations in object appearance and correlations in classifier responses, and accounting for localization uncertainty (Fig. 1). We confirm in MATLAB simulation that our method provides robustness with respect to the above sources of uncertainty compared to current methods. An ongoing work, initial simulations in a 3D Unreal Engine environment confirm that localization bias introduces class aliasing, causing wrong classification when uncertainty is not accounted for (Fig. 3).

While [4] limits itself to classification of a single object, more interesting challenges arise in a realistic scenario of an environment containing multiple objects, of which some may be instances of the same class, and with the general assumption that objects may be present for which classification

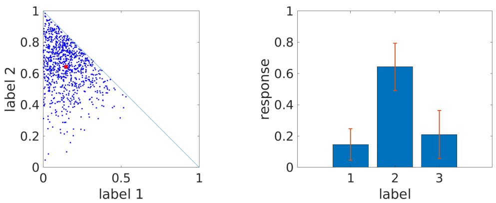



**Fig. 2:** Simulation of model uncertainty in classification to 3 classes. **Left:** multiple forward passes with MC dropout [6] result in a point cloud of classification outputs in the simplex, red point denotes sample mean. **Right:** Mean and standard deviation of classification scores over point cloud.

models are not available. Maintaining uncertainty as part of the semantic world map (possibly represented as a hybrid belief over discrete class variables and continuous poses and landmarks) may be of help in detecting and treating such novel information, possibly directing active collection of data for further training and disambiguation.

## REFERENCES


[1] N. Atanasov, B. Sankaran, J.L. Ny, G. J. Pappas, and K. Daniilidis. Nonmyopic view planning for active object classification and pose estimation. *IEEE Trans. Robotics*, 30:1078–1090, 2014.

[2] Israel Becerra, Luis M Valentín-Coronado, Rafael Murrieta-Cid, and Jean-Claude Latombe. Reliable confirmation of an object identity by a mobile robot: A mixed appearance/localization-driven motion approach. *Intl. J. of Robotics Research*, 35(10):1207–1233, 2016.

[3] C.M. Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag, Secaucus, NJ, USA, 2006.

[4] Yuri Feldman and Vadim Indelman. Bayesian viewpoint-dependent robust classification under model and localization uncertainty. In *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2018.

[5] Yarin Gal and Zoubin Ghahramani. Bayesian convolutional neural networks with bernoulli approximate variational inference. *arXiv preprint arXiv:1506.02158*, 2016.

[6] Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *Intl. Conf. on Machine Learning (ICML)*, 2016.

[7] Corina Gurău, Dushyant Rao, Chi Hay Tong, and Ingmar Posner. Learn from experience: probabilistic prediction of perception performance to avoid failure. *The International Journal of Robotics Research*, page 0278364917730603, 2017.

[8] G.A. Hollinger, U. Mitra, and G.S. Sukhatme. Active classification: Theory and application to underwater inspection. In *Robotics Research*, pages 95–110. Springer, 2017.

[9] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Advances in Neural Information Processing Systems (NIPS)*, pages 5580–5590, 2017.

[10] Beipeng Mu, Shih-Yuan Liu, Liam Paull, John Leonard, and Jonathan How. Slam with objects using a nonparametric pose graph. In *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2016.

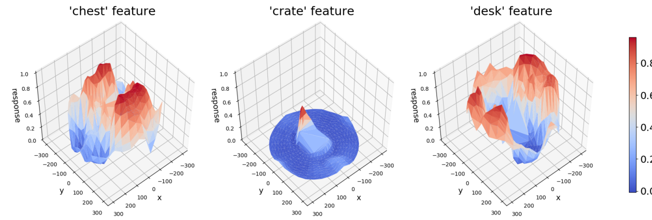
[11] Pavel Myshkov and Simon Julier. Posterior distribution analysis for bayesian inference in neural networks. *Advances in Neural Information Processing Systems (NIPS)*, 2016.

[12] Shayegan Omidshafiei, Brett T Lopez, Jonathan P How, and John Vian. Hierarchical bayesian noise inference for robust real-time probabilistic object classification. *arXiv preprint arXiv:1605.01042*, 2016.

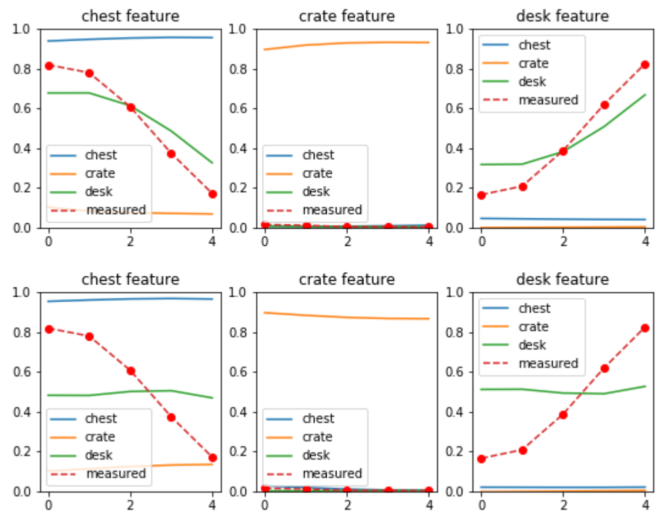
[13] T. Patten, M. Zillich, R. Fitch, M. Vincze, and S. Sukkarieh. Viewpoint evaluation for online 3-d active object classification. *IEEE Robotics and Automation Letters (RA-L)*, 1(1):73–81, January 2016.

[14] Weichao Qiu, Fangwei Zhong, Yi Zhang, Siyuan Qiao, Zihao Xiao, Tae Soo Kim, and Yizhou Wang. Unrealcv: Virtual worlds for computer vision. In *Proceedings of the 2017 ACM on Multimedia Conference*, pages 1221–1224. ACM, 2017.

[15] Charles Richter and Nicholas Roy. Safe visual navigation via deep learning and novelty detection. In *Robotics: Science and Systems (RSS)*, 2017.


**(a)** Spatial (2D) model of classifier responses for class 'desk' (interpolation using GP). For some viewpoints, most likely class is 'chest', motivating the use of a model over raw classifier outputs.




**(b)** Plots of classifier responses measured over a 2D track for an object of class 'desk' (red) against classification models (model for class 'desk' in green, 'chest' - in blue, 'crate' - yellow). **Top:** measurements best match model for ground truth class ('desk'). **Bottom:** localization bias causes measurements to shift against spatial model. As a result, measurements over first part of track better match 'chest' model, leading to erroneous classification if localization uncertainty is not accounted for.

**Fig. 3:** Spatial model for a class is represented using a separate GP learned per feature (see [4] for details). Here, training data is obtained by capturing rendered images of an object of corresponding class (using Unreal Engine with UnrealCV plugin [14]), feeding them into a Caffenet classifier, then fitting GP models to the components of interest of produced classification vectors. Plots in (b) correspond to localization uncertainty scenario from [4].


[16] Amit Shaked and Lior Wolf. Improved stereo matching with constant highway networks and reflective confidence learning. *arXiv preprint arXiv:1701.00165*, 2016.

[17] WT Teacy, Simon J Julier, Renzo De Nardi, Alex Rogers, and Nicholas R Jennings. Observation modelling for vision-based target search by unmanned aerial vehicles. In *Intl. Conf. on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1607–1614, 2015.

[18] Fabio Tosi, Matteo Poggi, Alessio Tonioni, Luigi Di Stefano, and Stefano Mattoccia. Learning confidence measures in the wild. In *British Machine Vision Conf. (BMVC)*, volume 2, 2017.

[19] Javier Velez, Garrett Hemann, Albert S Huang, Ingmar Posner, and Nicholas Roy. Modelling observation correlations for active exploration and robust object detection. *J. of Artificial Intelligence Research*, 2012.