Research Proposal

Adam Kosiorek

Artificial Intelligence has always tantalized me. Long before understanding what computer science stands for I wondered why AI has not been achieved yet. To explore it further, I enrolled in BSc in Robotics at the best technical university in my country. Programming and software-oriented courses were my favourite and paired with my passion for vision and perception, which I have developed as a photojournalist, they steered me towards my first computer vision course. At that moment it became apparent that pursuing a PhD was one of the few ways that could get me closer to solving the AI challange.

To learn from the best, I joined Computer Vision Lab at Samsung R&D, where I implemented a Bag of Words-based pipeline for object classification. Consequently, I learned about the state-of-the-art in keypoint detection and feature description, which later helped me to appreciate the importance of learned low level features in Convolutional Neural Networks (CNNs). Before jumping into deep learning for object classification and image duplicate detection, I ported the project to Android and researched codebook generation within BoW. In my bachelor thesis I investigated how spatial information affects classification performance by implementing a similar system for Kinect-gathered point cloud classification.

My observation was that deep learning, although often superior to other approaches, requires huge computational resources and datasets. To mitigate this problem I went to study computational science at the Technical University of Munich. Numerical linear algebra and scientific computing, comprising the majority of the coursework, help enormously with optimization and implementation of algorithms. Currently, I am working with Rudolph Triebel of the Computer Vision group on the introspective capacity of neural networks. It is understood as the ability of a classifier to assess uncertainty of its prediction given the data [1]. It is a paramount problem in mobile robotics and medicine where wrong classification might lead to loss of life. Our intermediate results show that neural networks, augmented with additional layers and a novel cost function, can be jointly trained for classification and uncertainty estimation. The topic might expand into my master thesis.

Last summer I did an internship at Bloomberg in London, where I worked on fraud detection in financial transactions. The problem can be cast as unsupervised anomaly detection with further verification in a supervised setting. I learned how difficult it can be to introduce an innovative approach in a corporate environment.

In my doctoral research project I would like to focus on deep neural networks for reinforcement learning and recurrent neural networks, preferably within the Mobile Robotics Group. Following is one of the research problems that attracts me (I am also open to other problems). Suppose that a mobile robot moves through highly differentiated space with multiple scene types. High accuracy classification of hundreds of object classes requires computationally expensive approaches and huge labeled datasets [2]. Since some objects can be more likely to appear in a particular type of a scene, it might be possible to develop a recurrent CNN that

adapts to the changing environment. One way of accomplishing this is to have the filters of the CNN tuned online in an unsupervised way, in either autoencoder [3] or adversarial [4] setting. It would make the CNN more responsive to the recently seen object classes. It might lead to the decalibration of the final fully-connected layers of the network, which might be mitigated by simultaneous fine tuning on a (small) training set. Additionally, if the set of classes predicted by the classifier could be altered at run-time, it might be possible to achieve higher accuracy at lower computational cost. This could be implemented by having A — a recurrent CNN fitted to the most prevalent object classes in the training set and B — a second classifier able to predict whole groups of classes. At any given point in time A would predict a subset of all classes, while B would maintain a probability-weighted ranking of groups of classes. When a significant change in the ranking occurs, A could be refitted to the union of the most probable groups of classes.

I expect this project to take around 3 years. The first stage, lasting about 3 months, would be to develop a recurrent CNN for object classification in sequential data similar to [5]. The next stage would be focused on unsupervised tuning with supervised calibration of the recurrent CNN. The design of appropriate architectures, adjusting learning algorithms and merging unsupervised tuning with supervised calibration in a unified approach might take 12 to 15 months. Finally, implementing dynamically changing prediction classes requires solving a number of challenges. Refitting of the final fully-connected layers would have to be made fast enough to work in real-time, since the system should work on a mobile robot. An algorithm for scoring groups of classes and change of a scene type would have to be developed. The ultimate goal would be to cast it as a differentiable problem, so that object classes to predict could be learnt with gradient-based approach, similar to how memory is learnt in [6]. This phase might take around 15 months.

My passion for solving technical problems coupled with my demonstrated skills show that I am ready and extremely motivated to carry out research as a PhD student at the University of Oxford. I am confident that I will be able to contribute significantly to and benefit immensely from my stay in Oxford. Thank you for considering my application.

References

[1] H. Grimmett, R. Triebel, R. Paul, and I. Posner, "Introspective classification for robot perception," *The International Journal of Robotics Research*, 2015.

[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[3] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[4] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[5] M. S. Pavel, H. Schulz, and S. Behnke, "Recurrent convolutional neural networks for object-class segmentation of rgb-d video," in *Neural Networks (IJCNN), 2015 International Joint Conference on*, pp. 1–8, IEEE, 2015.

[6] S. Sukhbaatar, J. Weston, R. Fergus, *et al.*, "End-to-end memory networks," in *Advances in Neural Information Processing Systems*, pp. 2431–2439, 2015.