Statement of Purpose

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. I studied maths, physics and mechanics, material science and manufacturing technologies; programming and software-oriented courses were my favourite, however. Paired with my passion for vision and perception, which I have developed as a photojournalist, it steered me towards my first computer vision course. I immediately knew that this intersection of image processing and machine learning was the area I wanted to focus on. It became apparent that pursuing a PhD was one of the few ways that could get me closer to solving the AI challenge.

To learn from the best, I joined Computer Vision Lab at Samsung R&D in Warsaw for an internship. My project was to design and implement 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 automatically learned low level features in convolutional neural networks. I focused my research primarily on the visual codebook generation within Bag of Words. The final step was to optimize it and then port it to the Android platform. In my bachelor thesis I investigated how spatial information affects classification performance by implementing a similar system for Kinect-gathered point cloud classification. Afterwards, I continued my work at Samsung, where I turned to deep learning techniques for object classification and image duplicate detection.

My observation was that deep learning, although often superior to older approaches, required huge computational resources and datasets. To mitigate this problem I went to study computational science at the Technical University of Munich. With cutting-edge computer vision research groups it seemed a great choice. 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 Caner Hazirbas and Rudolph Triebel of the Computer Vision group headed by Prof. Cremers. In our recent project we investigate 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. While high classification accuracy is desired, it is impossible to assess whether a given prediction is accurate in a test setting, where no labels are available. It is, therefore, vital to assess the uncertainty of predictions. 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.

This 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. This, together with my earlier industrial experience, convinced me that I do want to pursue a PhD. I love solving scientific problems, which do not have "the best" solution or a reference specification, by going into the deepest details.

In my doctoral research project I would like to focus on recurrent neural networks for computer vision. RNNs are well suited to sequential information processing and, while often used for NLP tasks, they have been somehwat underutilized in computer vision. If beliefs about content could be introduced in a form of priors inferred from previous elements of a sequence, it might be possible to increase object classification accuracy in videos and to shrink the network size. Another interesting problem is an end-to-end RNN for optical flow computation. The only end-to-end neural network, while efficient and accurate, does not use temporal information. It computes optical flow for any two possibly unrelated images [2]. I would like to investigate RNN designs capable of computing optical flow while presented with a single image at a time. Solving both problems might thus lead to smaller networks and reduced amounts of computation, possibly increasing energy efficiency of such systems.

I believe that the CS PhD programme at the University of Toronto is perfect for me due to the following reasons. Firstly, the faculty has an extensive research experience in neural networks for computer vision, natural language and multi-modal settings.

Firstly, the courses taught by world renown scientists e.g. Prof. Raquel Urtasun and Prof. Richard Zemel "Probabilistic Learning and Reasoning" provide an extraordinary introduction to research in deep learning I would like to conduct. Secondly, research of the Machine Learning group at the U of T is multimodal in that it successfuly merges natural language processing with computer vision [3]. I find it very relevant, for it might lead to the improvement of living standard for people with impaired sight.

My passion for solving technical problems coupled with my demonstrated skills show that I am ready and extremely motivated to pursue a PhD programme at the Stanford University. I am confident that I will be able to contribute significantly to and benefit immensely from my stay at Stanford. 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] P. Fischer, A. Dosovitskiy, E. Ilg, P. Häusser, C. Hazırbaş, V. Golkov, P. van der Smagt, D. Cremers, and T. Brox, "Flownet: Learning optical flow with convolutional networks," *IEEE International Conference on Computer Vision*, 2015.

[3] K. Xu, J. Ba, R. Kiros, A. Courville, R. Salakhutdinov, R. Zemel, and Y. Bengio, "Show, attend and tell: Neural image caption generation with visual attention," *arXiv preprint arXiv:1502.03044*, 2015.