

Title: Convolutional Neural Networks in Visual Object Tracking

Abstract: In 2012, convolutional neural networks (CNN) has been popularized through AlexNet architecture, which has won the ImageNet Large Scale Visual Recognition Challenge. Since then, CNN has been applied in various applications such as object tracking, visual surveillance, action recognition and many more. For single object tracking, MDNET tracker which uses CNN as the basis for feature extraction has won the visual object tracking (VOT) challenge in 2015. In this talk, I will explain the evolution of CNN-based tracker starting from DeepTrack until the most recent ones. There are three main components in devising a CNN-based tracker which are model initialization, search and matching procedures and model update. Model initialization function is to train the network parameters to fit the appearance model of the tracked object. Search and matching procedures then take the trained model and find the most similar patches in the consequent frames. Appearance model of the tracked object will normally change when the object of interest moves. Thus, the trained model needs to be updated and pruned to avoid model drift issue. However, the most important aspect of a CNN-based tracker is the network architecture itself. The most used architecture consists of several layers of CNN coupled with few fully connected (FC) layers. However, the majority of trackers use fixed pre-trained CNN layers, while the FC layers are trained online during the first frame. Few trackers also manage a collection of FC models instead of a single model, which are updated sequentially in various types of structure that include tree and hierarchical approaches. I will also share several other CNN-

based architectures of single object tracker. To train a CNN-based tracker, both positive and negative samples are collected by pivoting the sampling centroid on the tracked object. The same sampling procedures are also used in the model update, where the samples are collected for a few frames before the appearance model is retrained by initializing the parameters using the best-matched parameters.



Assoc. Prof. Mohd Asyraf Zulkifley

I received my Bachelor Degree of Engineering (International Islamic University Malaysia) and Ph.D. (The University of Melbourne) in 2008 and 2012, respectively. I have been working at the Universiti Kebangsaan Malaysia (The National University of Malaysia) since the start of my carrier, where I am currently attached to the Center for Integrated Systems Engineering and Advanced Technologies (Integra) as an associate professor. In 2014, I became visiting academic at the University of Melbourne, working with Prof. Bill Moran on robust visual tracking. In 2015, I have also become a visiting researcher for one year at MIMOS, which is the leading research center for ICT in Malaysia. Then, I did a two-year (2016-2018) attachment at the University of Oxford as a sponsored researcher working on deep learning-based visual object tracking under the guidance of Prof. Niki Trigoni. Since then, my main research interest has shifted towards deep learning

applications in computer vision problems. Up until August 2019, I have supervised and co-supervised six Ph.D. students until their completion. In term of research funding, I have managed to secure 10 research grants as principal investigator and another 15 research grants as co-researcher that amounted to more than RM 2 million. I am also an active reviewer for international journals such as IEEE Transaction on Pattern Recognition and Image Analysis, IEEE Signal Processing Letters, IEEE Access, Computer in Biology and Medicine, Computer Methods and Programs in Biomedicine and many more.