TRINITY COLLEGE DUBLIN

Abstract

Computer Science Department
School of Computer Science & Statistics

Master in Computer Science

Wireless Augmented Reality

Bringing deep learning to mobile augmented reality systems through computational offloading to the cloud

by Darragh McCaffrey

There is currently a huge amount of interest in bringing smart augmented reality applications to mobile devices. Mobile devices however, have limited resources and this means they cannot run certain computationally expensive computer vision techniques in a sufficiently fast manner. Taking the use case of object detection, the current state of the art object detection models utilise deep learning techniques to make accurate predictions. Deep learning incurs high processing and memory loads which results in long response time and a large amount of battery drain in mobile devices with limited resources. One solution to this problem is to offload these deep learning computations to the cloud. Offloading to a more powerful backend server provides the mobile application with increased processing power and reduces battery consumption but is subject to network impairments.

In this dissertation, the use of a reduced deep-learning model run locally on a mobile device is compared to offloading the computation to a much larger, more accurate deep learning model for object detection on the cloud. The impact of varying networking conditions on the performance of computational offloading to the cloud are also investigated and impairments are masked through local processing. This is completed through the development of a mobile cloud computing framework.

The evaluation of this framework shows a large increase in frame rate and accuracy and decrease in energy consumption when offloading. It was also shown that it is vitally important for the application to be contextually aware when making decisions on whether to offload or not. It was found that there are trade-offs between energy consumption, accuracy and speed when making this decision. Parameters such as bandwidth, network latency, image size and deep learning model size were shown to have a large impact on this decision of whether it is more favourable to perform detections locally or offload to the cloud.