THE UNIVERSITY OF HONG KONG

Faculty of Science

HKU-TCL Joint Research Centre for Artificial Intelligence

Principal Investigator (PI): Dr. Ngai WONG
Department of Electrical and Electronic Engineering
Faculty of Engineering

Project Title: A Fast and Lightweight Instance Segmentation Neural Network for Real-time Video Scene Understanding

Abstract:

Various object detection and segmentation applications have witnessed remarkable progress in recent years due to advances in deep neural networks (DNNs). The deep learning-based approaches can efficiently extract scene information using semantic segmentation. Nevertheless, pixel-wise approaches are designed to segment all pixels in a frame, thus incurring unnecessary computational complexity and high latency.

Proposal-based segmentation avoids handling all pixels by learning only the proposed object candidates. However, it still requires multiple rounds of expensive candidate proposals, and a large amount of segmentation time is wasted on the unadopted or overlapped areas of candidates. Moreover, most existing methods do not consider the temporal relationship of objects (viz., activities) in a video stream, which still remains a great challenge especially for its implementation on resource-constrained edge/terminal devices.

Lately, there are attempts to produce a single-stage image/video segmentation. YOLACT combines the prototype masks and predicted coefficients, and then crops with a segmented bounding box. PolarMask introduces the polar representation to formulate pixel-wise instance segmentation as a distance regression problem. And SOLO divides a network into two branches to generate instance segmentation with predicted object locations. Nonetheless, these methods still require significant pre- or post-processing before or after localization, and cannot offer a real-time speed. Furthermore, existing video detection and segmentation approaches are mostly based on image data, e.g., using CNNs to extract features from each time frame. With a huge amount of high dimensional video frame input, these methods suffer from high computational complexity and limited accuracy, hence cannot be fit onto resource-restrictive edge devices. A fast yet lightweight single-shot segmentation network is desired for real-time (or near real-time) video segmentation and understanding running on edge devices. Subsequently, this project proposes a scene understanding network by Single-Shot Segmentation, called S3-Net, for real-time video analysis and comprehension. There objectives are threefold:

1. To perform video analytics (specifically, video scene classification) targeting lightweight, board-level implementation.
2. To deploy S3-Net and/or its variant(s) on AI development boards.
3. To explore research-valuable, innovative ways of video classification.