Reimplementing SiamRPN[1] for Single Object Tracking

EECS504 Final Project Proposal

Group Member

Junjiang Ye; Danny Chen; Ziyu Guo

Introduction

Visual object tracking is an essential foundation in various tasks of computer vision, such as automatic driving, industries, and video surveillance. SiamRPN is a high-performance algorithm used for single object tracking problem. We are trying to reimplement SiamRPN and test in multiple datasets.

Problem Statement

It is challenging to accomplish visual object tracking due to the large appearance variance caused by camera motion, illumination, deformation, and occlusion. In order to solve the problem practically in high-speed FPS, we will fully reimplement SiamRPN on Pytorch, and use different datasets to train and test the model, then evaluate its performance and robustness in the end.

Proposed Solution

Overall, SiamRPN is composed two parts: Siamese Network and Region Proposal Network. Siamese Network extracts the features from the input datasets, and Region Proposal Network predicts the location and confidence of the target. The dimensions of these two parts are raised by convolution operations so that they are unified in an end-to-end framework.

First, we will implement Siamese Network by using two Alexnets. One is template branch whose input is template frame, another is detection branch whose input is detection frame. The weight parameters of two Alexnets are exactly the same.

Second, we will implement Region Proposal Network. After receiving their features, we first implement depth-wise convolution in each channel, they send them to Classification Branch and Regression Branch to get the predictions of class and location.

Third, we will use different datasets, such as OTB, VOT and so on, to train our model.

Fourth, we will use one-shot detection to implement inference. For the first frame, the template branch will extract two features from the input, and these two features will be transformed to two convolution layers and put in the Classification Branch and Regression Branch after detection branch; for the subsequent frames, they will go into this new RPN and finally get results of classification and location. After comparing all the weighted scores of boxes, we can choose the highest one as final result.

Evaluation Method

We aim to build a workable model which can classify and track the object at a reasonable success rate and can run in acceptable FPS speed. We will first apply the model to typical scenarios, and then calculate the corresponding indicators of the model to evaluate its overall performance. The potential evaluation metrics include Precision, Average Overlap Rate, Robustness, Area under Curve and FPS, etc.

---

[1] Li, Bo, et al. "High performance visual tracking with siamese region proposal network." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.