Siamese Trackers

Group Members: Liming Gao, Chengyu Hu, Zekai Liu, Mingzhao Yu

Date: 02/09/2020

## The baseline tracker

Many algorithms have been proposed for visual object tracking in recent years. SiamFC first considered tracking to be a similarity learning problem[1]. Due to its simple network structure, high speed, and good accuracy, there are many follow-up studies, including the Siamese-RPN. Its recourse code in MATLAB and Pytorch is available for SiamFC. Thus, we are going to use the SiamFC as our baseline tracker.

## The datasets we are going to use

We're going to use VOT and OTB as our main datasets. The specific version is not decided yet, but we believe that since the paper we're looking at in our project mainly used these two datasets, we can prove that we solve the drawbacks better if we use the same group of datasets. We may also use UAV20L to see the performance on longer videos. All the datasets mentioned above have a relatively small number of videos which allows our machine to perform better.

## The problem we are going to fix

Siam-FC only tracking offline to use very limited training data to build a tracker rather than online learning which cannot adapt to various appearance changes. Siam-FC trackers highly rely on similarity scores to predict the target’s position. In this way, there are no good solutions for some complex tracking scenarios including motion estimation, fast motion, occlusion, shape deformation which caused by changing the selection method of the search area for real-time application.

Siam-FC only performs a simple correlation operation to obtain the position of the current frame target during the tracking process, which makes it very dependent on the cosine window that adds to the score map to filter the background interference around the target in the tracking process. For instance, the gap between the targets’ score and background’s score in perturbed texture will become so small that the cosine window will have a major impact on the final score and mislead the tracker. The target will be further away from the center of the cosine window.

## The approach

We propose to introduce a Kalman filtering method to make a stable prediction of the target’s motion trajectory, and the target search area in the tracking process is tailored according to the target position obtained by Kalman filtering.

## Reference

[1] Bertinetto, L.; Valmadre, J.; Henriques, J.F.; Vedaldi, A.; Torr, P.H. Fully-convolutional siamese networks for object tracking. In Computer Vision—ECCV 2016 Workshops, Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016; Springer: Cham, Switzerland, 2016; pp. 850–865.