Long-Term Tracking

Radoslav Atanasoski

I. Introduction

In this assignment, a deep CNN-based tracker called SiamFC is set up and tested on a new dataset containing longer sequences then from the previous assignments. Furthermore, the tracker is upgraded with a re-detection function to re-detect the object that is being tracked after occlusions or when the tracker goes off the object. Both approaches, with and without re-detection are tested on the entire dataset and compared using Precision, Recall and F-score. Also, two different approaches are used for randomly sampling regions for re-detection and different number of sampled regions per frame.

II. Experiments

A short-term tracker SiamFC

In this task, the short-term tracker SiamFC is set up and tested on the provided dataset. The tracker is short-term, as it doesn't incorporate re-detection when the tracker goes of the object its tracking. We can see the performance of the tracker on the dataset in table II. We can see that on the majority of the sequences the tracker tracks poorly. The reason is that the sequences contain occlusions, where the target disappears behind an object, and reappears in a few frames at a different location. Without the re-detection function, the tracker fails to find the object and tracks something else in the background, resulting in poor performance for long-term tracking sequences.

Sequence	Precision	Recall	F-score
car9	0.636	0.270	0.379
cat1	0.840	0.415	0.555
deer	0.718	0.200	0.312
dog	0.204	0.073	0.108
person14	0.670	0.038	0.073
person20	0.735	0.735	0.735
sitcom	0.545	0.504	0.524
skiing	0.336	0.044	0.078
sup	0.542	0.476	0.507
Total	**0.607**	**0.299**	**0.400**

Table I: Results of testing short-term SiamFC.

Figure 1: Example of using short-term SiamFC on car9.

In figure 1, we can see tracking results using the initial short-term SiamFC, without the re-detection function. We can see that, when the object is occluded, the tracker starts learning on the occlusion and never tracks the target again. From this, can see the importance of using re-detection for long-term tracking.

Long-term tracker

In this task, the short-term tracker SiamFC is upgraded for long-term tracking, by adding a re-detection function. When the maximum response score falls bellow a certain threshold, the re-detection stage begins, where the tracker tries to re-detect the object using N random samples for each frame.

Sequence	Precision	Recall	F-score
car9	0.602	0.561	0.581
cat1	0.772	0.538	0.634
deer	0.633	0.183	0.284
dog	0.063	0.052	0.057
person14	0.690	0.630	0.658
person20	0.735	0.735	0.735
sitcom	0.489	0.507	0.498
skiing	0.556	0.240	0.335
sup	0.520	0.476	0.497
Total	**0.586**	**0.414**	**0.486**

Table II: Results of testing long-term SiamFC.

The results are obtained with a threshold of 4 and activating the re-detection after the response score being lower then the threshold for 10 frames. Also, a scaling factor of 1.2 is used when searching for the object, and only 1 Uniform random sample per frame. We can see the clear difference in the importance of the re-detection for long-term tracking. Although in most sequences, the tracker re-detected the target, there are a few (dog, deer, skiing) in which the re-detection failed to find the target.

Figure 2: Example of using long-term SiamFC on car9.

In figure 2, we can see that in the first image, the car is being tracker, while on the second and third image, the car is occluded and the tracker is in the re-detection phase, using Uniform random sampling to find the car. Finally, in the forth image, we see the car re-detected.

Testing on different thresholds

For the confidence score, the maximum of the responses is used to indicate if the tracker should go into tracking or re-detection mode. The threshold value heavily determines the success of the tracking. If the threshold value is too low, the tracker will re-detect a background object and will falsely track. If the threshold is too high, the tracker will deem the target that is being tracker, to not be the true target, and constantly enter into re-detection phase. Thus, determining the optimal threshold value is very important. In table IV, we can see results of the long-term SiamFC on the entire dataset, using different threshold values.

Threshold	Precision	Recall	F-score
3	0.571	0.349	0.433
3.5	0.634	0.384	0.478
4	**0.598**	**0.435**	**0.504**
4.5	0.579	0.397	0.471

Table III: Results of testing long-term SiamFC using different threshold values.

We can see that the tracker performs best (highest F-score) when the threshold value is around 4. Thus, throughout the experiments, this threshold value is used.

Testing using multiple random samples per frame

In table IV, we can see that by using more random samples per frame, does help the tracker find the target faster. However, the trade-off comes with the price of speed. Using a significantly higher number of multiple samples can slow down the tracker.

N samples	Precision	Recall	F-score
5	0.594	0.421	0.493
10	0.603	0.401	0.481
20	0.550	0.407	0.468
50	0.595	0.416	0.490
100	**0.595**	**0.442**	**0.507**

Table IV: Results of testing long-term SiamFC using multiple random samples per frame.

Testing with different sampling methods

The choice of sampling methods plays an important role in the re-detection phase. Initially the Uniform random sampling methods was used throughout the experiments. However, in table V, we can see results of using a Uniform random sampling and Gaussian random sampling. Furthermore, we can combine both sampling methods. For example, in the first 100 detections, the Gaussian random sampling can be used, searching around the location where the target was last detected. After that, if the target is still not detected, the Uniform random sampling method is used, as there is higher chance of the target repairing somewhere in the distance from its initial disappearance. Furthermore, the sigma for the Gaussian can be increased every frame while the tracker is in the re-detection phase.

Method	Precision	Recall	F-score
Uniform	0.567	0.392	0.463
Gaussian (fixed)	0.609	0.319	0.419
Gaussian (growing)	**0.597**	**0.413**	**0.488**
Gaussian (fixed) + Uniform	0.573	0.393	0.466
Gaussian (growing) + Uniform	0.586	0.416	0.487

Table V: Results of testing long-term SiamFC using different sampling methods.

We can see that the Uniform random sampling method outperforms the Gaussian (fixed) by a fair margin. It makes sense, as in the majority of the sequences, when the target disappears, it reappears somewhere further away from its initial position of disappearance. On those sequences, the tracker completely fails as it randomly samples around the location when the target was last being tracked. However, if sigma is increased, the tracker improves, as it samples from a larger area which can be seen in the Gaussian (growing). The sigma is increased every frame the target is not detected, greatly improving the re-detection capability.

III. Conclusion

In this assignment, tha SiamFC is set up at tested on a provided dataset. Furthermore, the SiamFC is upgraded for long-term tracking by adding a re-detection mechanism. Both short-term and long-term trackers are compared on the entire dataset, of which we can see the importance of re-detection for long-term tracking.

Two different sampling techniques are explored, the Uniform and Gaussian. The Gaussian is tested wth a fixed sigma and then with a growing sigma, of which the Gaussian with the growing sigma shows significant improvement.

Furthermore, the long-term tracker is tested using different threshold values and different number of random samples per frame, of which a threshold value with around 4 gives best results and higher number of samples shows to improve re-detection speed.