Long-term tracking

Blaž Erzar

I. Introduction

In this exercise, we compare two approaches for tracking, short-term and long-term tracking. For the short-term tracking, we use the already-implemented SiamFC tracker. For the long-term tracking, we extend this tracker by implementing the re-detection procedure. Both trackers are evaluated on the provided long-term dataset.

II. Experiments

A. SiamFC tracker

The SiamFC tracker is based on the discriminative correlation filter. The long-term tracker starts re-detection when the confidence (explained in the next subsection) falls below the threshold of 4. During re-detection, 20 search regions are sampled uniformly and re-detection is stopped if the target is found in one of them.

The results of the evaluation of both trackers is shown in Table I. We can see that precision drops slightly, so some of the re-detections are not optimal. But, the recall and consequently the F_1 score do increase, meaning the overlap is larger when the target is visible.

Tracker	**Precision**	**Recall**	**F_1 score**
Short-term	0.62	0.31	0.41
Long-term	0.58	0.45	0.51

Table I: Evaluation results for the short-term and long-term SiamFC trackers on the whole long-term dataset.

B. Confidence score

For the confidence score, we use the maximum peak height of the DCF response. When the confidence score falls below the threshold, the re-detection procedure starts. It is stopped when the target patch score is above the threshold. Looking at the tracker performance for different threshold values in Figure 1, we can see that the F_1 score is the highest for thresholds between 4 and 5. We used threshold 4 in all evaluations.

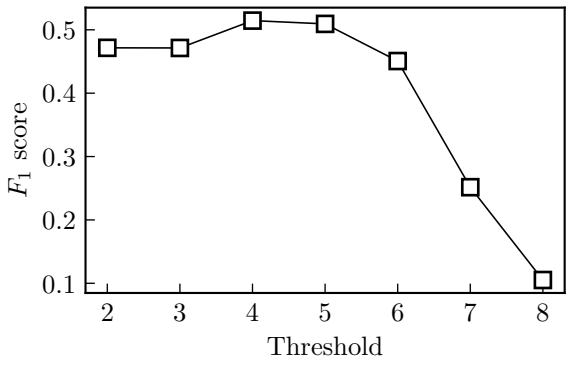

Figure 1: Long-term tracker performance on the whole dataset for different re-detection threshold values.

C. Number of samples

The number of samples for search regions during re-detection affects the tracker speed as well as re-detection performance. From Figure 2, we can see that for most sequences, re-detection is faster with more samples (darker lines).

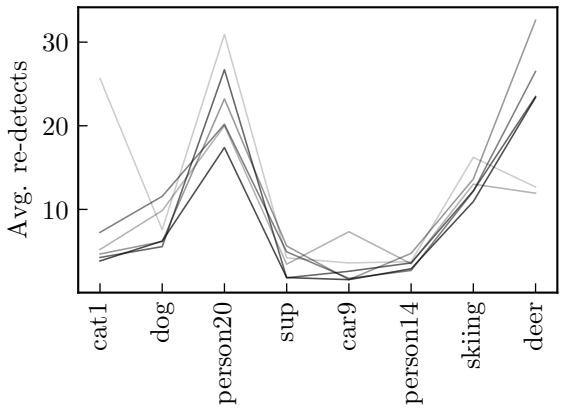

Figure 2: Average re-detection lengths for each sequence in the dataset for different number of samples. The number of samples increases in increments of 5, from 5 for the lightest line, to 30 for the darkest one.

In some cases, this is not the case. This is probably due to the fact that sampling is probabilistic.

D. Sampling

As the simples sampling approach, we first implemented the uniform sampling of search regions. Another method we use is Gaussian sampling, where we increase its variance in each re-detection frame. The comparison of both samplings is shown in Table II.

Sampling	**Precision**	**Recall**	**F_1 score**
Uniform	0.58	0.45	0.51
Gaussian	0.58	0.40	0.47

Table II: Evaluation results for the long-term tracker using uniform and Gaussian search region sampling and 20 samples in each frame.

In some cases, the Gaussian sampling may work better, but the variance should be tuned per sequence. In our case, the uniform sampling works better. The search regions cover the majority of the frame and work better when the target goes off screen and appears at a different location.

E. Re-detection

We visualize the tracking results in Figure 3 and search region samples obtained uniformly. We can see that the tracker successfully detects the occlusion and when the target is visible again, one of the search regions does include it. In this case, the tracker successfully re-detects its location.

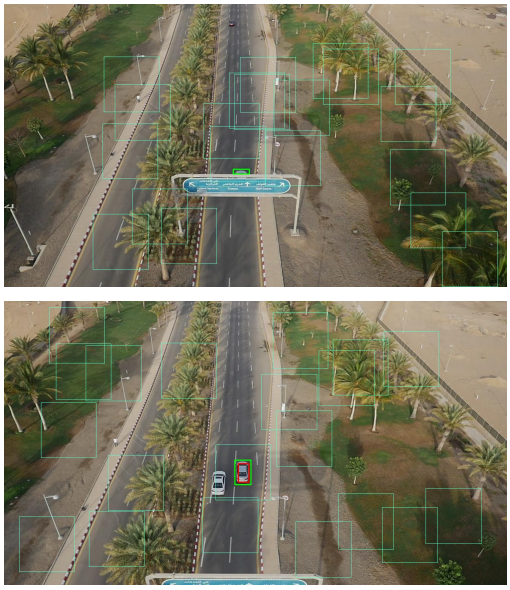

Figure 3: Tracker visualization for the car9 sequence. The ground truth bounding box is shown in green, tracker prediction in red and re-detection samples in light blue. In the top image, the target is occluded and the tracker does not make a prediction. In the bottom image, the tracker successfully re-detects the target in one of the search region samples.

The search region sizes are the as for the short-term SiamFC. When the best search region is found, the tracker is ran again to find the optimal bounding box scale for the target.

III. Conclusion

We have shown that a short-term tracker can be modified into a long-term tracker to increase its performance on a long-term dataset. To detect occlusions, we use a simple confidence score and the best threshold selected using a simple line search. The number of samples affects the re-detection length and we found that in most cases, generating more samples is benefitial. In the end, we also try the Gaussian sampling instead of the uniform one, but it did not yield any improvement.