Exercise 5: Long-Term Tracking

Ana Poklukar

I. Introduction

In this exercise, we extended the SiamFC tracker [1] for long-term object tracking by adding failure detection and re-detection capabilities. The original SiamFC, designed for short-term tracking, struggles with occlusions and target disappearances. We addressed this by introducing a confidence-based mechanism to detect tracking failures and a re-detection strategy based on candidate sampling and similarity scoring. We evaluated both the original and modified trackers on a subset of the long-term VOT benchmark dataset [2], and analyzed the impact of different re-detection parameters on tracking performance and robustness.

II. Experiments

Firstly, we set up and executed a deep convolutional neural network (CNN)-based tracker, SiamFC, which is a fully-convolutional Siamese network designed for short-term object tracking. We evaluated the performance of the SiamFC tracker on the benchmark dataset.

To extend SiamFC for long-term tracking, we modified it to detect target disappearances and enable re-detection. This was achieved by introducing a threshold on the maximum correlation response—when the response dropped below this value, the tracker inferred the target was lost. For re-detection, we randomly sampled candidate positions across the image and evaluated them using the Siamese network, selecting the one with the highest similarity score. This allowed the tracker to recover when the target reappeared after occlusion or leaving the frame.

Table I presents the results of the initial evaluation of both the original short-term SiamFC tracker and the modified long-term version, tested on the dataset mentioned previously. These results highlight the limitations of the original SiamFC model in handling long-term tracking scenarios, particularly in cases of target disappearance or occlusion. For the long-term tracker, we used 30 samples in the re-detection sampling strategy and set the tracking uncertainty threshold to 3. Compared to the baseline, the long-term tracker achieved improvements in recall and F-score, demonstrating a better ability to recover from tracking failures. Although the precision slightly decreased, the overall performance benefited from the added re-detection capability, resulting in a more robust and complete tracking trajectory.

Table I
Performance comparison between the baseline short-term SiamFC tracker and the modified long-term version, evaluated on the full dataset.

Tracker	**Precision**	**Recall**	**F-Score**
SiamFC	0.602	0.299	0.399
Long-Term SiamFC	0.592	0.398	0.476

Figures 1 and 2 illustrate a comparison between the short-term and long-term SiamFC trackers on the challenging `car9` sequence. In Figure 1, the short-term tracker fails to recover after an occlusion event, losing the target entirely—this is evident as the red box (tracker output) diverges from the green ground truth box. In contrast, Figure 2 demonstrates the effectiveness of the modified long-term tracker. After the same occlusion, it successfully re-detects the target shortly afterward. The final detection (large green box) aligns well with the ground truth (small green box), and the red boxes represent the 30 candidate regions sampled during re-detection using the uniform strategy.

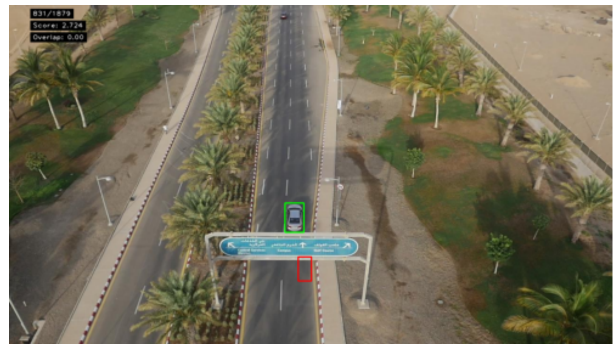

Figure 1. Failure case of the short-term SiamFC tracker on the `car9` sequence.

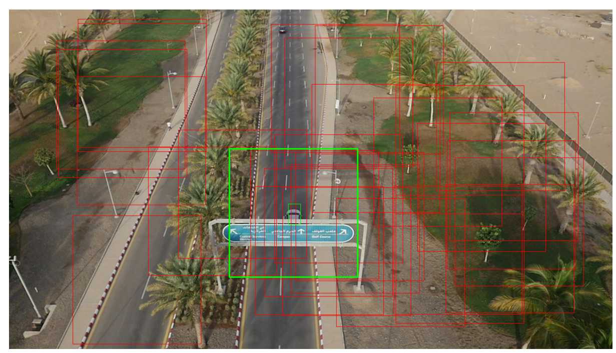

Figure 2. Successful re-detection by the long-term SiamFC tracker on the `car9` sequence.

Tracking Failure Detection via Confidence Score

In our implementation, we used simple thresholding to determine whether the object was lost. To define the confidence score, we first calculate a reliability score q_t for each frame t as:

$$q_t = \max(R_t) \times PSR(R_t),$$

where R_t is the response map of the tracker, and PSR is the Peak-to-Sidelobe Ratio. This score captures both the strength and sharpness of the correlation response peak, indicating the tracker's certainty about the target's location.

The Peak-to-Sidelobe Ratio is a commonly used measure in correlation-based tracking to quantify the distinctiveness of the response peak. It is computed as the difference between the maximum value in the response map and the mean of the sidelobe region, divided by the standard deviation of that sidelobe region:

$$PSR = \frac{\max(R_t) - \mu_{sl}}{\sigma_{sl}}.$$

A high PSR indicates that the peak stands out clearly from the surrounding values—implying a confident match—while a low PSR suggests ambiguity or noise in the tracking response.

The confidence score c_t is then defined as:

$$c_t = \frac{\overline{q}_t}{q_t},$$

where $\overline{q}_t$ is the mean of q_t over all past frames. A high value of c_t suggests a drop in current confidence relative to historical levels, signaling a potential tracking failure.

Through experimentation, we set the threshold value τ_q to 3. When $c_t > \tau_q$, we infer that the target has likely been lost, and we trigger the re-detection module. Lowering the threshold would make the tracker more sensitive to changes in the object's appearance, causing re-detection to occur more frequently.

Effect of Sample Count on Re-detection Performance

We then evaluated how varying the number of randomly sampled candidate regions during re-detection affects the performance of the long-term tracker.

As shown in Table II, increasing the number of samples generally improves recall and F-score up to a certain point, with the highest F-score observed at 50 samples. This indicates that a moderate increase in sampling density enhances the likelihood of correctly locating the target after a disappearance, without overly compromising precision. However, further increasing the sample count to 100 yields diminishing returns and slightly lower recall, possibly due to increased noise or redundant sampling.

Table II
Impact of the number of randomly sampled regions during re-detection on the performance of the long-term SiamFC tracker.

Samples	Precision	Recall	F-Score
10	0.615	0.402	0.486
30	0.592	0.398	0.476
50	0.602	0.436	0.506
70	0.584	0.403	0.477
100	0.604	0.399	0.480

In terms of re-detection speed, using more samples tends to reduce the number of frames needed to re-detect the target, since the probability of sampling a region near the true target location increases. However, this comes at the cost of higher computational load, which can impact runtime performance. A sample count of 50 appears to offer a good balance between re-detection accuracy and efficiency.

Sampling Strategies for Re-detection

Next, we explored how different sampling strategies affect re-detection performance in the long-term SiamFC tracker. Specifically, we compared uniform sampling—where candidate regions are drawn uniformly across the entire image—to Gaussian sampling centered around the last confident position, using a fixed standard deviation $\sigma = 1$.

As shown in Table III, Gaussian sampling led to a higher precision, suggesting that it generates fewer false positives by focusing the search near the most probable target location. However, this improvement in precision comes at the cost of a lower recall, likely because the narrower search area may miss the target if it has moved significantly from the last known position. Consequently, the F-score is slightly lower for Gaussian sampling, indicating a trade-off between search focus and recovery capability. Overall, uniform sampling offers better robustness in cases where the target may reappear farther from the last known location, while Gaussian sampling is beneficial in scenarios with more spatial continuity.

Table III
Comparison of sampling strategies for candidate region selection during re-detection in the long-term SiamFC tracker.

Sampling	Precision	Recall	F-Score
Uniform	0.592	0.398	0.476
Gaussian	0.626	0.359	0.456

III. Conclusion

In this exercise, we extended the SiamFC tracker to handle long-term tracking scenarios by incorporating failure detection and re-detection mechanisms. Our modified tracker demonstrated improved performance over the original version, particularly in sequences with occlusions or target disappearances. We showed that the confidence-based thresholding method effectively identifies tracking failures, and that re-detection using randomly sampled candidates enables recovery. Additionally, we analyzed the effects of the number of re-detection samples and different sampling strategies, highlighting the trade-offs between precision, recall, and computational cost. Overall, our approach improves robustness in long-term tracking while maintaining competitive accuracy.

References

[1] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr, "Fully-convolutional siamese networks for object tracking," 2021. [Online]. Available: https://arxiv.org/abs/1606.09549

[2] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. Čehovin Zajc, T. Vojir, G. Bhat, A. Lukežič, A. Eldesokey, G. Fernandez, and et al., "The sixth visual object tracking vot2018 challenge results," 2018, subset of the VOT2018-LT dataset downloaded from http://box.vicos.si/alanl/dataset-lt.zip.