Long-term Tracking

Stefanela Stevanović, 63220492

I. Introduction

In this project, we explored the performance of the SiamFC tracker both as a short-term and long-term tracker. We evaluated its precision, recall, and F-score on car9 sequence. To evaluate the re-detection capability, we experimented with different numbers of randomly sampled regions and measured the number of frames required for successful re-detection. Finally, we explored different sampling techniques during target re-detection, such as Gaussian sampling around the last confident position with a fixed and growing standard deviation. We compared the results of these sampling methods to the uniform sampling used in the original SiamFC tracker, analyzing how they affect the tracker's performance.

II. Experiments

A. Short-term vs. Long-term Tracking

To transform the SiamFC tracker into a long-term tracker, we introduce modifications that improve the tracker's ability to re-detect the target.

- Confidence Score Computation: A confidence score, which indicates the tracker's confidence in the presence of the target at the computed position, was defined as the maximum response divided by the template response computed at the first frame (when object was being tracked). If the tracker is failing to track the object, the confidence score is set to 0.
- Thresholding for Target Loss: To detect target loss or failure, the maximum correlation response is thresholded. If the response falls below a predefined threshold, it indicates that the tracker is likely unable to locate the target accurately. A threshold value of 3.5 has been found to work the best through experimentation.
- Thresholding for Re-detection: To initiate re-detection after target loss, a separate threshold value was determined. This threshold was dynamically updated every frame when the target was visible. The threshold was set based on the mean of previous responses minus 0.25. This ensures that the threshold remains adaptive to changes in the tracking scenario and facilitates successful re-detection after target loss.
- Re-detection Strategies: When target loss occurs, the tracker triggers the re-detection process to locate the target in subsequent frames. In the first experiment we performed the random sampling with 50 samples over the entire image to generate candidate regions for potential target re-detection. Random seed for the sampling was always set to the number of the current frame.

The presented table (Table I) displays the performance results of both the short-term and long-term versions of the SiamFC tracker tested on car9 sequence. The performance is evaluated using three metrics: precision, recall, and F-score.

TABLE I: Short-term and Long-term SiamFC tracking performance with 50 randomly sampled regions.

Tracker	Precision	Recall	F-score
SiamFC Short-term	0.64	0.27	0.38
SiamFC Long-term	0.60	0.59	0.59

The short-term tracker achieves a precision of 0.64, indicating that approximately 64% of the tracked regions are correctly classified as the target object, while the long-term tracker has slightly lower precision of 0.60. Recall, which indicates the percentage of the actual target object regions that are successfully tracked is significantly greater for the long-term SiamFC, 59% comapred to 27%. Overall, the results suggest that the long-term tracker performs better than the the short-term tracker, given that the F-score for the long-term tracker is higher, indicating a better balance between precision and recall compared to the short-term tracker.

B. Impact of the different number of randomly sampled regions on the re-detection capability

In the second experiment to evaluate how different numbers of randomly sampled regions during re-detection impact the re-detection capability of the tracker. The objective was to determine whether the number of sampled regions influences the time taken to re-detect the target. We conducted the experiment with 15, 25, 50 and 100 randomly sampled regions and the results are shown in table II.

TABLE II: Impact of different numbers of randomly sampled regions on re-detection capability in terms of number of the frames taken to re-detect the target.

Number of samples	Number of frames for re-detection
15	73
25	54
50	50
100	38

From the result in table II, we can see that increasing the number of samples has a positive impact on re-detection. Higher numbers of sampled regions enables the tracker to explore a larger portion of the space, leading to the less number of frames needed for the target re-detection.

C. Tracking visualization

We choose to visualize the result from the SiamFC tracker with 100 random samples, that takes 38 frames to re-detect lost target on car9 sequence. Tracking results are show on Figure 1 and Figure 2.

Fig. 1: Frame 790. Target is lost due to occlusion

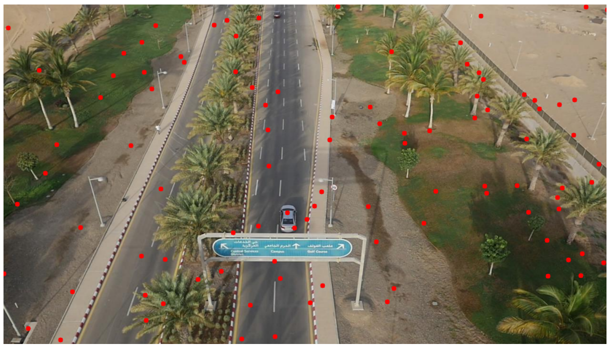

Fig. 2: Frame 828. Target is found. Red dots represent 100 random samples

D. Gaussian sampling during target re-detection

In this experiment we implemented Gaussian sampling with 50 samples around the last confident position in the re-detection process. We implemented both Gaussian sampling with fixed standard deviation and Gaussian sampling with a growing standard deviation. For Gaussian sampling with fixed standard deviation, we tried the following standard deviations: 5, 20, 35, 50, 65 and 80. For Gaussian sampling with growing standard deviation we started with the specified initial standard deviation and increased the standard deviation by 1 in each new iteration (in each frame where we try to re-detect the target). We tried initial standard deviations of 5, 20 and 35. The results of the experiment are shown in table III and table IV.

TABLE III: Impact of different standard deviations in Gaussian sampling with fixed standard deviation on re-detection capability in terms of number of the frames taken to re-detect the target.

Standard Deviation	Number of frames for re-detection
1	
5	37
20	40
35	38
50	37
65	43
80	51

From the results presented in Table III, it is evident that the utilization of Gaussian sampling with fixed standard deviation (std) consistently outperformed uniform sampling with the same number of samples (50). The experiments revealed that the tracker required fewest frames to re-detect the target when employing Gaussian sampling with std values of 5 and 50. However, as the std increased to 65, the number of frames needed for re-detection began to rise. These findings suggest that Gaussian sampling with fixed std values enhances the re-detection capability of the tracker compared to uniform sampling. The improved performance can be attributed to several factors. Firstly, Gaussian sampling allows for a more focused exploration of the search space by concentrating the sampling points around the last confident position. This increases the likelihood of capturing the target object within the sampled regions.

TABLE IV: Impact of different standard deviations in Gaussian sampling with growing standard deviation on re-detection capability in terms of number of the frames taken to re-detect the target.

Initial Standard Deviation	Number of frames for re-detection
1	40
5	48
20	43
35	50

Using a growing std in each iteration implies that the spread of the candidate regions gradually increases over time. While we didn't manage to achieve better result with Gaussian sampling with growing std starting from different initial standard deviations, this approach can be beneficial in scenarios where the target object undergoes significant changes in appearance or experiences occlusions. By gradually expanding the search area, the tracker can potentially locate the target even if it moves further away from its last known position.