VISUAL TRACKING VIA SIAMESE NETWORK WITH GLOBAL SIMILARITY

Chao Fan, Yulong Wang, Chenglong Li, Jin Tang

School of Computer Science and Technology, Anhui University, Hefei, China

BACKGROUND & MOTIVATION

- Many trackers based on Siamese network train their networks by utilizing either pairwise loss or triplet loss, which easily leads to over-fitting.
- It is difficult to distinguish some hard samples in the training samples.
- We propose a novel global similarity loss to train the siamese network.

OUR TRACKER

- The pipeline of our tracker

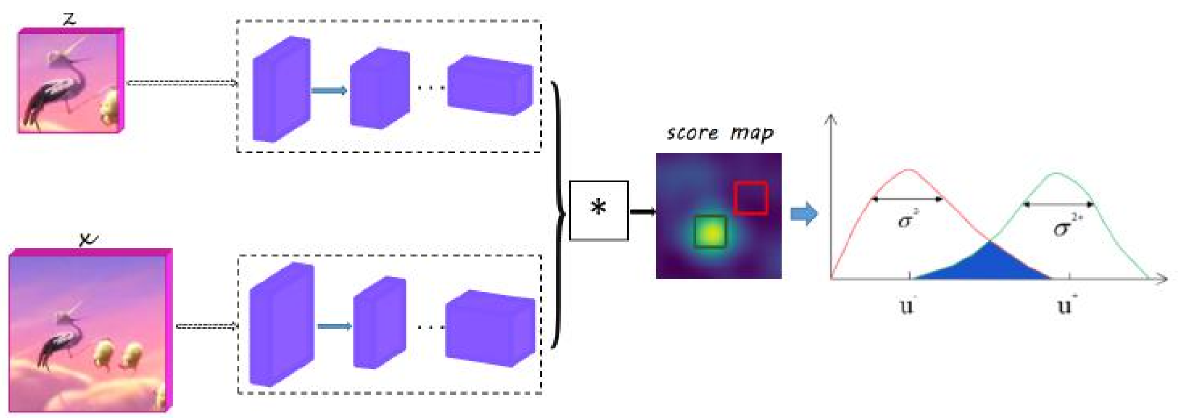


Fig. 1. Exemplar image z input from the upper branch and search image x input from the lower branch and the score map was obtained by cross-correlation. We use two Gaussian distributions to simulate the distribution of positive and negative sample similarity scores, the red curve represents the distribution of negative sample similarity scores, and the green curve represents the distribution of positive sample similarity scores. we set the mean and variance of these two distributions to be μ^+, μ^-, σ^{2+} and σ^{2-} respectively.

Global Similarity Loss

- we use two Gaussian distributions to simulate the similarity score map of positive and negative samples. The red curve represents the similarity score distribution of negative samples, and the green curve represents the similarity score distribution of positive samples.

$$l_{dis} = (\sigma^{2+} + \sigma^{2-}) + \lambda * max(0, \delta - (\mu^+ - \mu^-))$$

where σ^{2+} and μ^+ represent the variance and mean of the score distribution of positive sample similarity, σ^{2-} and μ^- represent the variance and mean of the score distribution of negative sample similarity, and δ is the margin between the mean of the positive and negative similarity score distributions. In the experiment, we set δ to 0.1 and λ is a hyper parameter that balances the importance of each term and we set it to 1.0.

Hard Sample Loss

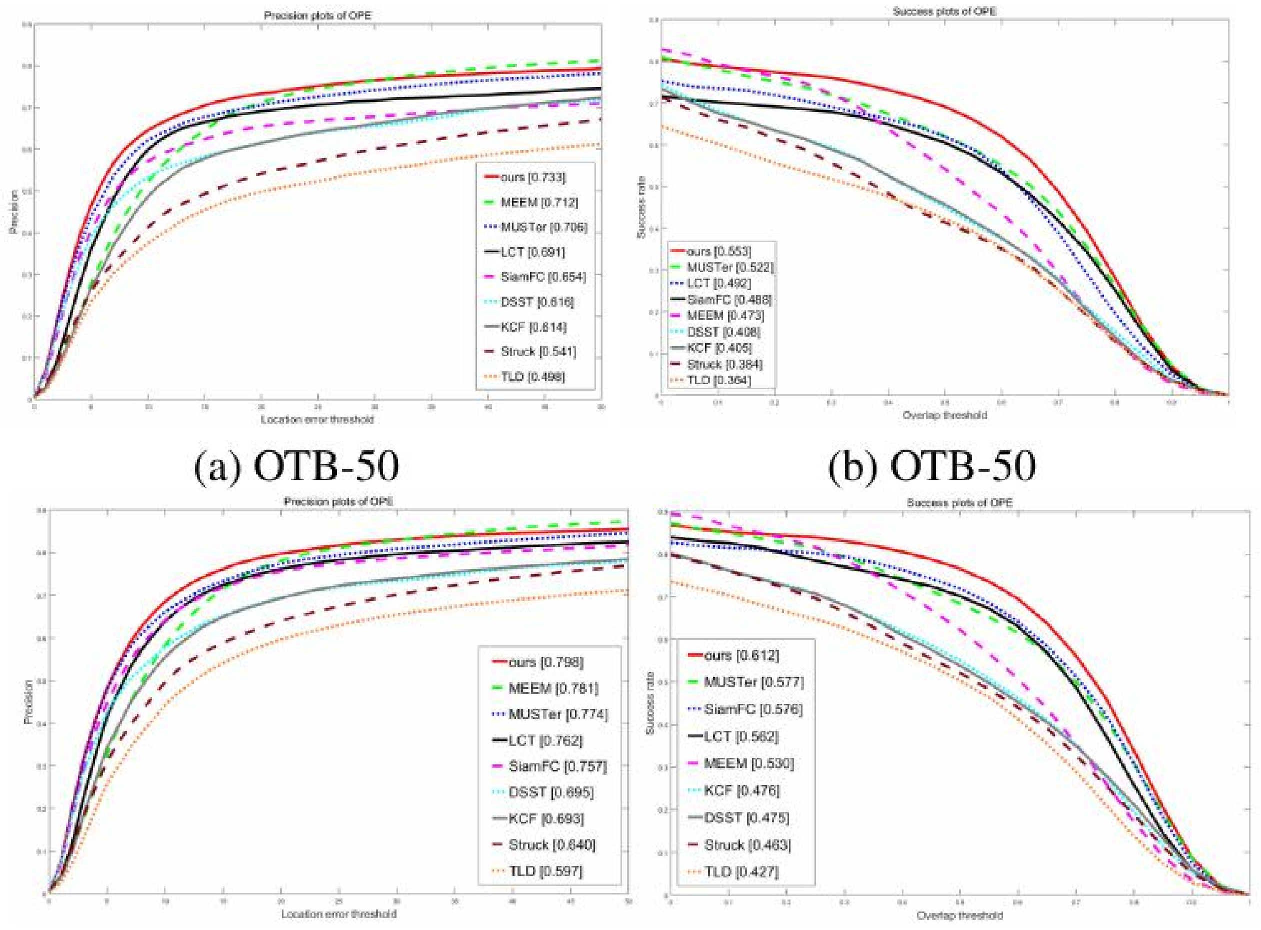
- There still be some samples that are difficult to distinguish in the edge part of the two distributions. We make the maximum similarity score of the negative sample less than the minimum similarity score of the positive sample by a margin, thus further reducing the area of the blue part.

$$l_{hard} = (max(neg) - min(pos) + \alpha)_+$$

- Through the above analysis, our total loss function is written as follows:

$$L_{total} = l_{dis} + \eta * l_{hard}$$

EXPERIMENTS



(a) OTB-50 (b) OTB-50

(c) OTB-100 (d) OTB-100

- The first row is the result of baseline, and the second row is the result of our method.