

Reviewer: Ling Zhang, February 6, 2020

Citation: Bo li, Wei Wu, Qiang Wang, Fangyi Zhang, Junliang Xing, Junjie Yan, “*SiamRPN++: Evolution of Siamese Visual Tracking with Very Deep Networks*”, CVPR2019.

Quick Summary: The authors aim to solve the problem about the network. AlexNet is widely used, however, when researchers change the AlexNet to other deeper networks such as ResNet-50, the performance becomes worse. The authors break the restriction through *a simple yet effective spatial aware sampling strategy*. They successfully train a ResNet-driven Siamese tracker with significant performance gain.

Ideas/Approach/Result:

Idea: Although the Siamese trackers have outstanding performance, especially for well balanced accuracy and speed, the accuracy still has a gap with the state-of-the-arts on tracking benchmarks. The authors they all use AlexNet. Then they try deeper network ResNet which expected perform better than AlexNet. However, it gain no performance. They find that the core reason comes from the destroy of the strict translation invariance. Other trackers use AlexNet with zero-padding, so they satisfies this invariance restriction.

The authors introduce *a simple effective sampling strategy* to break the spatial invariance restriction of the Siamese tracker. They successfully train a Siamese-RPN tracker using the ResNet as a backbone network and obtained significant performance.

Methods: Most of design is similar to the previous review paper “*High Performance Visual Tracking with Siamese Region Proposal Netwrok*”. The main difference is that *Siamese tracker* uses AlexNet, the *Siamese-RPN tracker* which trained in this paper uses ResNet (more deeper than AlexNet). Also, the authors use layer-wise aggregation. They use ResNet-50 and explore multi-level feature extracted from the last three residual block for layer-wise aggregation. The other part is the most important method in this paper: ***Depth-wise Cross Correlation*** which is the core operation to embed two branches information.

SiameseRPN use cross correlation embed much higher level information than SiamFC does by adding a UP-Xcorr. The heavy up-channel module make imbalance of parameter distribution which makes optimize training hard in SiamRPN. Compare with UP-Xcorr, the ***Depthwise Cross Correlation*** (DW-Xcorr) has more light weight. DW-Xcorr layer contain 10 times fewer parameters than UP-Xcorr used in SiameseRPN. The performance of these two correlation modules are similar. After replacing the cross correlation to depthwise correlation, the network have high reduce of computational cost and the memory usage. Also, the training procedure is more stable because the template and search branches’ parameter balanced.

Result: The performance of their SiameseRPN++ tracker improved. Compare with other state-of-the-arts trackers, SiameseRPN++ tracker obtain the comparable performance in OTB-2015 dataset, VOT-2018 dataset, VOT2018 Long-term dataset, UAV123 dataset, LaSot dataset, and TrackingNet dataset.

Compare SiameseRPN++ with SiameseRPN: SiameseRPN++ uses lower computational cost and memory usage than SiameseRPN. In the experiment part, we can find the performance in each dataset challenge, SiameseRPN++ all get better accuracy and speed than SiameseRPN. The use of DW-Xcorr is not only lower the resource usage, but also balance the parameter in template and search branch which make the training optimization easier.