

SiamMask++: More Accurate Object Tracking through Layer Wise Aggregation in Visual Object Tracking

Authors : Hyunbin Choi, Jihyeon Noh, Changwon Lim

Abstract : In this paper, we propose SiamMask++, an architecture that performs layer-wise aggregation and depth-wise cross-correlation and introduce multi-RPN module and multi-MASK module to improve EAO (Expected Average Overlap), a representative performance evaluation metric for Visual Object Tracking (VOT) challenge. The proposed architecture, SiamMask++, has two versions, namely, bi_SiamMask++, which satisfies the real time (56fps) on systems equipped with GPUs (Titan XP), and rf_SiamMask++, which combines mask refinement modules for EAO improvements. Tests are performed on VOT2016, VOT2018 and VOT2019, the representative datasets of Visual Object Tracking tasks labeled as rotated bounding boxes. SiamMask++ perform better than SiamMask on all the three datasets tested. SiamMask++ is achieved performance of 62.6% accuracy, 26.2% robustness and 39.8% EAO, especially on the VOT2018 dataset. Compared to SiamMask, this is an improvement of 4.18%, 37.17%, 23.99%, respectively. In addition, we do an experimental in-depth analysis of how much the introduction of features and multi modules extracted from the backbone affects the performance of our model in the VOT task.

Keywords : visual object tracking, video, deep learning, layer wise aggregation, Siamese network

Conference Title : ICCV 2022 : International Conference on Computer Vision

Conference Location : Vancouver, Canada

Conference Dates : September 22-23, 2022