Accuracy and Long-Term Tracking via Overlap Maximization Integrated with Motion Continuity

Wenhua Zhang, Haoran Wang, Zhongjian Huang, Yuxuan Li, Jinliu Zhou, Licheng Jiao
School of Artificial Intelligence, Xidian University
Xi'an, Shaanxi Province, 710071, China
zhangwenhua_nuc@163.com lchjiao@mail.xidian.edu.cn

Abstract

The baseline is ATOM which aims at solving the problem of accurate target state estimation by proposing a novel tracking architecture. The architecture consists of dedicated target estimation and classification components. Classification component is trained online to guarantee high discriminative power in the presence of distractors. Target estimation is performed by the IoU-predictor network inspired by the IoU-Net which was recently proposed for object detection as an alternative to typical anchor-based bounding box regression techniques. In this work, we further enhance the performance of ATOM by embedding Squeeze-and-Excitation (SE) blocks into IoU-Net in ATOM to recalibrate useful features and suppress useless features and obtain ATOMFR. To solve the abnormal changes in the target box in ATOMFR, we add the Relocation Module on ATOMFR and get ATOMFR (RL). To solve the occlusion problem, we introduce the Inference Module into ATOMFR (RL) and obtain ATOMFR (RL + InF). Experimental results on VisDrone2019-SOT test set demonstrate the state-of-the-art performance of ATOMFR (RL + InF) compared with several existed trackers and it ranks the second place among all competitors.

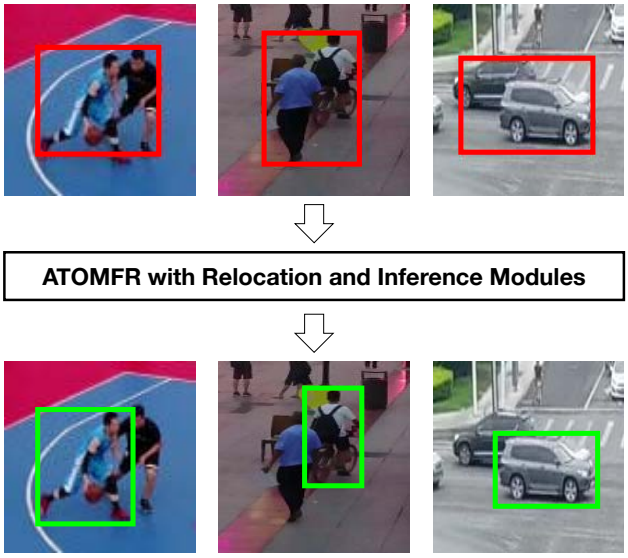

Figure 1. Tracking results between ATOM and ATOMFR (RL + InF). The top line is the tracking result of ATOM and the bottom line is the tracking result of ATOMFR (RL + InF). The frames are from sequences *uav0000087_00290_s, uav0000073_00038_s and uav0000229_00600_s* in VisDrone2019-SOT test set respectively.

1. Introduction

Object tracking, as a fundamental task in computer vision, has a wide range of applications, such as transportation surveillance, smart city, human-computer interaction and so on [41]. In this work, we focus on a sub-task of object tracking-single object tracking (SOT) on drone-based data set. SOT aims to estimate the state of a object, indicated in the first frame, across frames in an online manner [16, 6, 35, 20, 41, 19, 36, 7, 28]. Although researchers have made considerable progress in several natural data sets (OTB [38], VOT [27, 18], TempleColor [22]) [40, 20, 36, 19, 7, 5, 2, 5, 28, 34, 32, 14, 11, 1, 6, 21, 4, 26, 25, 30, 39, 8, 10], there are still deficiencies in using these methods directly on drone-based data sets such as view point changes and scale variance.

Discriminative Correlation Filter (DCF) based tracking methods as an important branch in object tracking, has attracted more and more attention due to its high tracking speed and performance. The core idea of DCF is to train a filter by minimizing a least-square loss. Due to DCF [16] use cyclic shifts to select samples, the correlation operation can simplify the calculation in the Fourier domain to achieve high tracking speed. In addition, there are many DCF based works aiming to improve performance, such as those apply multi-dimensional features [25], part-based strategies [23] and end-to-end learning [32]. In addition to

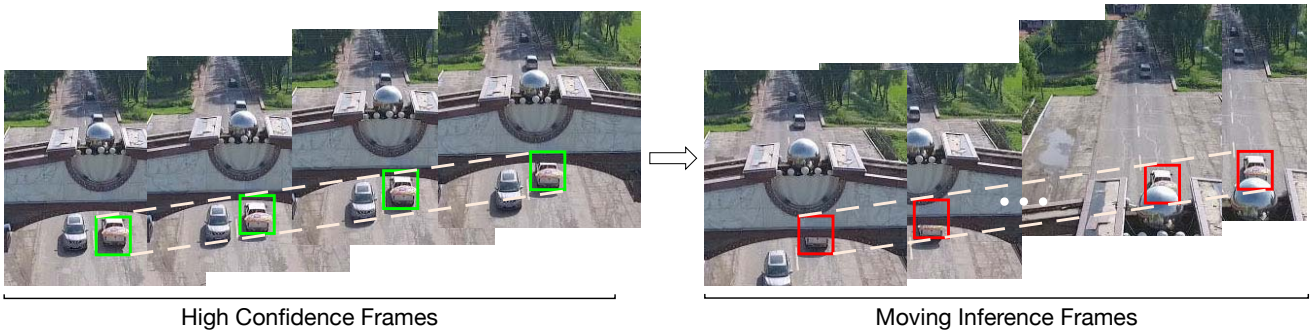

Figure 2. Moving inference process under occlusion situation.

tracking speed, DCF based trackers usually online train the filters and perform well on object discrimination. Expect for DCF, RPN based trackers ([20, 36, 19]) are also an important branch in object tracking. Take SiamRPN [20] as an example, SiamRPN consists of a template branch and a detection branch. RPN perform multi-scale proposal extraction on the correlation feature maps. In general, the RPN based trackers achieves good performance in object locating due to extensions offline training. Drawing on these two advantages, ATOM [7] was proposed as a two-stage tracking framework. ATOM divides the target tracking into two networks: target estimation network and the target classification network. The former is used for coarsely locating the target and the latter is used for fine locating the target. The target estimation network trains the IoU-Net [17] offline based on a large number of data sets. The target classification network uses a deep regression network with two convolutional layers for online tracking.

ATOM overcomes two difficulties existing in preceding trackers. One is that RPN based trackers are usually not accurate enough in object discrimination. Another is that DCF based trackers are general lack of bounding box accurate locating capabilities. But the performance of ATOM tracking is largely determined by IoU-Net, especially when there is an interference target in the search box, the final bounding box of IoU-Net will be too large to affect the performance (Fig. 1). Therefore, it is necessary to improve the performance of IoU-Net. Moreover, the occlusion problem is also evident in the drone data, which is prone to the problem of tracking lost. Therefore, it is necessary to achieve continuous tracking in this case.

Squeeze-and-Excitation (SE) Networks [15] demonstrate that if the importance of each feature channel is automatically obtained through the learning to enhance the useful features and suppress the features that are not useful for the current task and then the extracted features are more effective for representing the object. In this paper, to improve the performance of IoU-Net in ATOM, we use SE-ResNet blocks to replace the ResNet blocks in IoU-Net to recalibrate more useful features and name the new method ATOMFR. To further improve the scale adaptive ability in ATOMFR, we introduce the Relocation Module (Sec. 4.2) in ATOMFR, and obtain ATOMFR (RL). To deal with the problem of occlusion, we assume that the target moves at a constant speed for a short period. Then we infer the positions of the frames that are poorly differentiated based on the positions of the well-located frames by introducing the Inference Module as shown in Fig. 2 (Sec. 4.3). We call the new tracker ATOMFR (RL + InF). Experimental results on VisDrone2019-SOT test set demonstrate the state-of-the-art performance of ATOMFR (RL + InF) compared with all other competitors, and the superior performance of ATOMFR (RL + InF) among existed trackers.

2. Related Work

Before describing the proposed method, we first briefly introduce two types of methods that are most relevant to the proposed method: DCF based trackers and RPN based trackers.

DCF based Trackers. Target classification network in ATOMFR is constructed by training powerful discriminatively classifiers. Recently, DCF based trackers have achieved wide popularity. The first DCF based tracker MOSSE [3] was proposed by online training a filter via minimizing the output sum of squared error. Then DCF based trackers have been widely researched. In [16], the tracker was proposed by exploiting the circulant structure of the training samples and training the filter with HOG features in a kernel space. The CSR-DCF [24] proposed in [9] builds DCF with channel and spatial reliability. C-COT [12] employs a continuous-domain formulation and ECO [6] as its enhanced version which improves both speed and performance by several efficient strategies. MCCT [35] constructs multiple experts based on DCF and the suitable expert is selected with a robustness evaluation strategy.

RPN based Trackers. Another type of trackers that is closely related to our tracker is based on RPN. RPN [31] was first proposed in object detection field to extract multi-

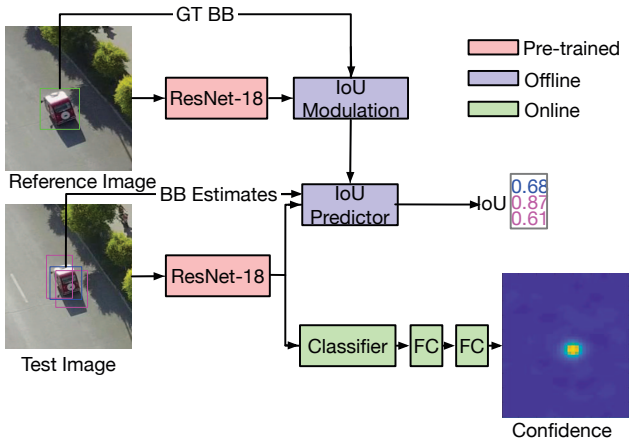

Figure 3. The framework of ATOM.

scale object candidate boxes. Subsequently, RPN was introduced to the object tracking field. SiamRPN [20] is a classic RPN based tracker. The SiamRPN includes a Siamese subnetwork for feature extraction and a region proposal subnetwork, where the region generation network includes two branches of classification and regression. SiamRPN++ [19] adopts a deeper convolutional network and uses some targeted strategies to further improve the performance of RPN based trackers. Naturally, SiamRPN and its extensions [19, 36] have advantages in object location due to extensive offline training. However, RPN based trackers is usually embedded in the Siamese based framework. This type of trackers lack online training and the performance of the target classification is usually not good enough. Similar to ATOM, ATOMFR (RL + InF) learns the classifier online and utilize extensive offline training for the target estimation task (Sec. 4).

3. Preview of ATOM

Nowadays, DCF based tracking methods and RPN based tracking methods are two important branches in SOT field. In general, the existing DCF based trackers lack flexible scale adaptability and the existing RPN based tracking methods lack sufficient discriminative power. In order to balance the trade-off between scale flexibility and discriminative power respectively in these two types of trackers, Danelljan $et.al$ proposed ATOM. Fig. 3 shows the structure of ATOM. ATOM divides the target tracking into two networks: target estimation network and target classification network. The former is used for coarse locating and the latter is used for fine locating. The target estimation network uses the IoU-Net [17], and the target classification network uses a DCF based classifier in a fully convolutional manner. The following mainly introduces these two networks.

Target Estimation Network. ATOM uses IoU-Net as the target estimation network (Fig. 4). IoU-Net is trained to predict the IoU between an image object and input bounding box candidate. In details, given a deep feature representation (through backbone network (ResNet-18 [13])) of an image x ($x \in R^{W \times H \times D}$), and an estimated bounding box B of an image object, IoU-Net predicts the IoU between B and the object. Here $B = (\frac{c_x}{w}, \frac{c_y}{h}, log_w, log_h)$, where (c_x, c_y) denote center coordinate of the bounding box and (w, h) represents the size of the bounding box. PrPool [17] is adopted to pool the region in x given by B. Therefore, the feature map x_B is of a predetermined size.

Target Classification Network. The target classification network used in ATOM is a fully convolutional neural network with two fully convolutional layers. This network is defined as follows:

$$f(x;\omega) = \phi_2(\omega_2 * \phi_1(\omega_1 * x)) \quad (1)$$

where x is the feature map output through the backbone network, ω_1, ω_2 are the filters in this network, ϕ_1, ϕ_2 are activation functions and $*$ represents the standard convolution operator. The loss function of the network is defined as follows:

$$L(\omega) = \sum_{j=1}^{m} \gamma_j ||fx_j;\omega - y_j||^2 + \sum_k \lambda_k ||\omega_k||^2 \quad (2)$$

where y_j is the label for each training samples. As in DCF trackers, y_j is a sampled Gaussian function centered at the target location. The importance of each training sample is indicated by γ_j. The weight of the filter ω_k is denoted by λ_k. This optimization strategy employed in this network is a Conjugate-Gradient-based strategy.

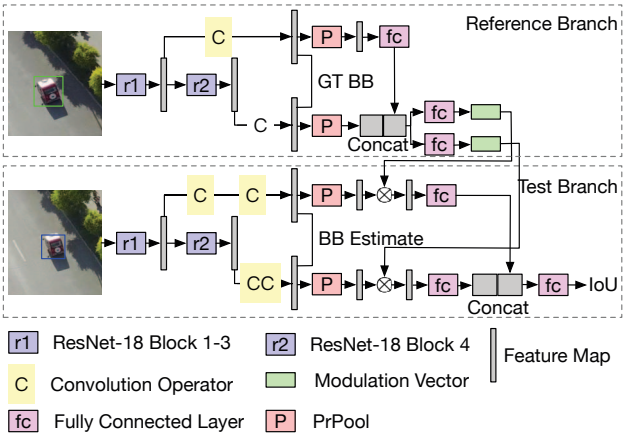

Figure 4. The framework of the target estimation network in ATOM.

4. Proposed Method

SE-Net [15] has demonstrated that the feature channels play important roles in object feature representation. In-

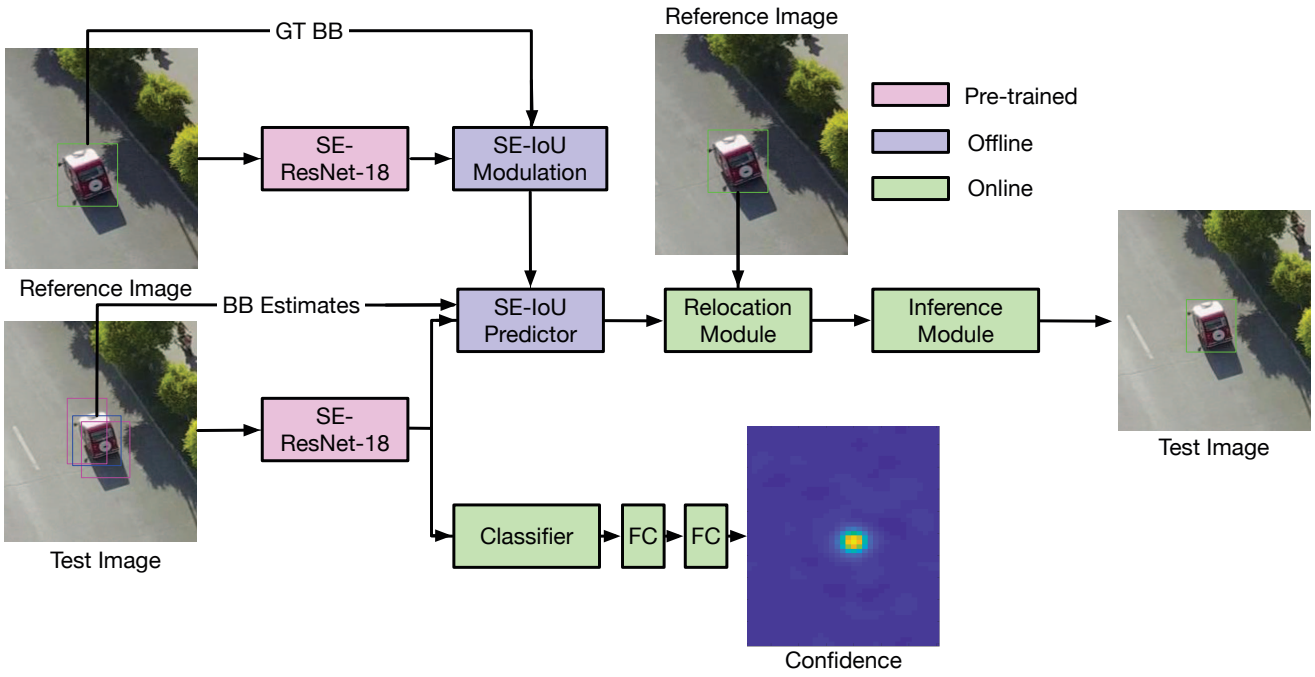

Figure 5. The framework of the proposed tracker ATOMFR (RL + InF).

spired by the idea of SE-Net, this paper continues to use the IoU-Net (SE-IoU-Net) as the target estimate network, and uses SE-ResNet-18 as the backbone network to adaptively learn the importance of each feature channels (ATOMFR) (Sec. 4.1). Naturally, this choice can improve the performance of IoU-Net by enhancing the representation of useful features. Considering that DCF based tracking methods have advantage in object classification, this paper continues to use DCF based method to complete the classification task. Namely, the target classification network is exactly the same to that of ATOM. In order to further improve the performance (scale adaptive ability) of ATOMFR, we introduce the Relocation Module (Sec. 4.2) into ATOMFR and obtain ATOMFR (RL). To solve the phenomenon of tracking loss when the occlusion occurs, we assume that the object moves at a constant speed in a short period, and then propose the Inference Module (Sec. 4.3) and construct ATOMFR (RL + InF). Fig. 5 shows the overall structure of the proposed tracker (ATOMFR (RL + InF)).

4.1. Feature Recalibration Module

The SE-Net [15] can adaptively select features that are useful to represent the object while suppress features that are useless to represent the object. The SE blocks embedded into ResNeXt [37], Inception-ResNet [33] and ResNet [13] have verified the performance of the SE blocks. In order to improve the performance of ATOM, this paper replace ResNet blocks with SE-ResNet blocks in IoU-Net in ATOM to enhance the useful features of the object while suppress useless features. Fig. 6 shows the structures of SE-ResNet module and ResNet module. Based on the ordinary residual block, SE-ResNet module introduces additional five layers, sequentially, global pooling layer, full connected layer, ReLU layer, full connected layer and sigmoid layer.

The global pooling (Global Pooling Layer) operation $O_{gp}(\boldsymbol{u}_c)$ is defined as follows:

$$O_{gp}(\boldsymbol{u}_c) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} u_c(i,j) \tag{3}$$

where u_c donates the c-th feature map with size $H \times W$. These two fully connected (FC) layers are used to adjust dimension of the output of global pooling layer through nonlinearity with reduction ratio r (in our experiment, $r = 16$). ReLU is the activation function that activates the output of the first FC layer. The final output $\boldsymbol{X}$ of the SE block is obtained by rescaling $\boldsymbol{U}(\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{u}_2, ..., \boldsymbol{u}_c])$ with the output feature maps $S(S = [s_1, s_2, ..., s_c])$ of the second FC layer activated by Sigmoid by Eq. (4).

$$\boldsymbol{X} = O_{se} = \boldsymbol{u}_c . \boldsymbol{s}_c \tag{4}$$

where . donates channel-wise multiplication, $\boldsymbol{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_c]$ and s_c is of size $H \times W$.

We use the SE-ResNet blocks to replace ResNet blocks as the basic unit of the residual block, and also use the 18-layer structure (SE-IoU-Net) as the backbone network in

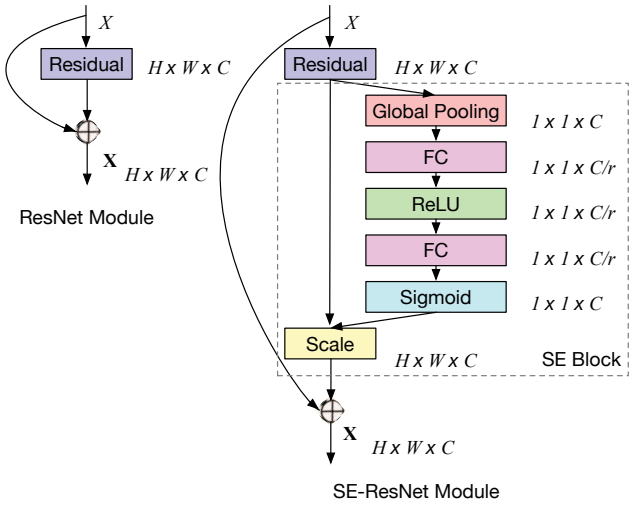

Figure 6. The structures of ResNet and SE-ResNet modules.

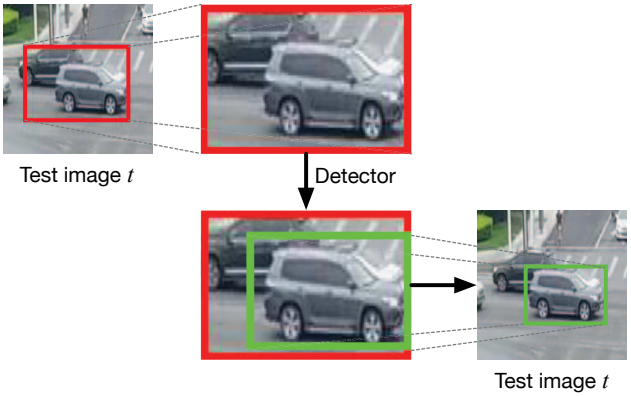

Figure 7. The framework of Relocation module.

ATOMFR. Similar to the IoU-Net in ATOM, the input of SE-IoU-Net is also composed of four parts: (1) features extracted through backbone network from current frame, (2) estimated bounding box (BB) in the current frame, (3) features extracted through backbone network from a reference frame, (4) the object bounding box in the reference frame. The outputs of SE-IoU-Net are IoU scores for each of the estimated bounding box (BB estimates) in current frame. The final bounding box is generated by maximizing the IoU score using gradient ascent in the tracking process.

4.2. Relocation Module

In order to further improve the performance of ATOMFR, the Relocation Module (a detector) is introduced to relocate the position of the tracked object which most likely inaccurate locating (ATOMFR (RL)). The detector adopts a two-stage detection framework, i.e. Faster RCNN with FPN (ResNet101 [13] as the backbone). Since DCF based trackers are sensitive to the presence or absence of the object in a frame (confidence score), the detector is fine-tuned according to the score of the classifier in the tracking process. That is, when the score of the DCF classifier is relatively high and the object is existed with a high confidence, then this frame is fine-tuned to the detector. When DCF classifier score is low, we don't use this frame to update parameters of the detector. When the object changes abnormally, we use the detector to redetect the object in this frame. In general, when the object bounding box is abnormal, IoU-Net score will be lower or the classifier's score will be lower. So in both cases, we use the detector to recorrect the tracking bounding box in the current frame. Fig. 7 shows the process of Relocation Module in ATOMFR.

4.3. Inference Module

Fig. 2 shows the inference process of the Inference Module. Considering that the VisDrone2019-SOT test data set is generally with long sequences, occasionally occlusion problem occurs. In order to improve the tracking efficiency under this condition, we assume that the motion of the object is uniform for a short period. P is used to indicate the location information of the object. P has four parameters, namely, $P \in \{x, y, w, h\}$, (x, y) refers to the coordinate of the top-left corner of the object and (w, h) donates size (width and height) of the object. Based on this assumption, if the location of the object in t-frame is P_t and the location of the object in $t+m$-frame is P_{t+m}, and then the location of the object in $t+m+q$-th frame (P_{t+m+q}) can be inferenced (Eq. (5)). Here m needs to satisfy $(m<\delta)$, where δ is a threshold (in our experiment, $\delta = 4$) and $(q = 1)$.

$$
\begin{aligned}
P_{t+m+q}(x_{t+m+q}) &= x_{t+m} + \frac{x_{t+m} - x_t}{m} \times q \\
P_{t+m+q}(y_{t+m+q}) &= y_{t+m} + \frac{y_{t+m} - y_t}{m} \times q \\
P_{t+m+q}(w_{t+m+q}) &= w_{t+m} + \frac{w_{t+m} - w_t}{m} \times q \\
P_{t+m+q}(h_{t+m+q}) &= h_{t+m} + \frac{h_{t+m} - h_t}{m} \times q
\end{aligned}
\tag{5}
$$

In detail, when the tracking score through ATOMFR (RL) in the $t+m+q$-frame is relatively low, we use the Inference Module (Eq. (5)) to infer the position of the tracked object in this frame.

Implementation Details. The training method and data sets (expect for VisDrone2019-SOT train set) of SE-IoU-Net in ATOMFR (RL + InF) is the same as that of IoU-Net in ATOM, and the parameters in the target classification network in ATOMFR (RL + InF) are the same as those in ATOM.

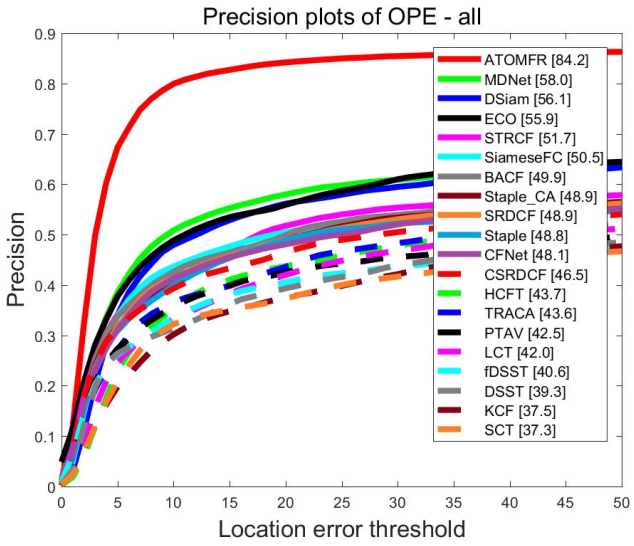

Figure 8. Precision plot on VisDrone2019-SOT test set with several existed methods.

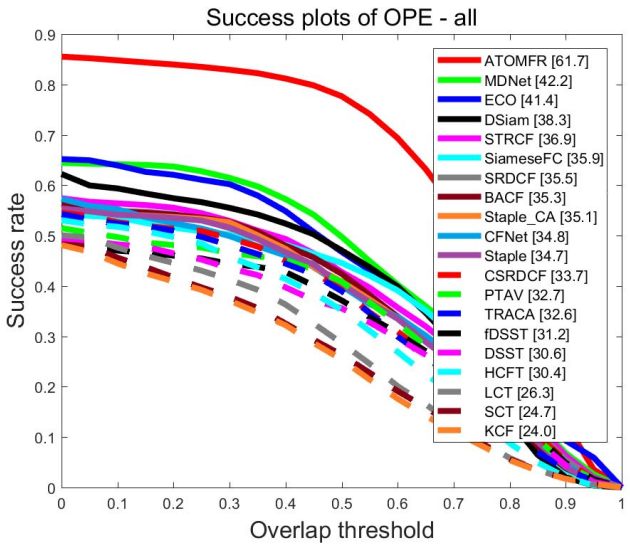

Figure 9. Success plot on on VisDrone2019-SOT test set with several existed trackers.

5. Experimental Verification

5.1. Experimental Setup

Data set. We use VisDrone2019-SOT test set [41] to evaluate the performance of the proposed tracker (ATOMFR (RL + InF)), other competitors' trackers and several existed trackers. This data set contains 60 sequences, which are *uav0000375_00001_s*, *uav0000231_03240_s*, *uav0000389_00001_s*, *uav0000023_00870_s*, *uav0000392_00001_s*, *uav0000393_00001_s*, *uav0000398_00001_s*, *uav0000372_00001_s*, *uav0000319_01840_s*, *uav0000229_00600_s*, *uav0000396_00001_s*, *uav0000388_00001_s*, *uav0000069_01200_s*, *uav0000397_00001_s*, *uav0000244_00479_s*, *uav0000079_00720_s*, *uav0000075_00088_s*, *uav0000387_00001_s*, *uav0000075_00240_s*, *uav0000378_00001_s*, *uav0000391_00001_s*, *uav0000328_04137_s*, *uav0000087_00290_s*, *uav0000328_01564_s*, *uav0000400_00001_s*, *uav0000191_00000_s*, *uav0000246_01416_s*, *uav0000073_00038_s*, *uav0000385_00001_s*, *uav0000394_00001_s*, *uav0000365_00001_s*, *uav0000077_00000_s*, *uav0000390_00001_s*, *uav0000311_02583_s*, *uav0000152_00750_s*, *uav0000181_00725_s*, *uav0000367_00001_s*, *uav0000374_00001_s*, *uav0000328_02760_s*, *uav0000382_00001_s*, *uav0000383_00001_s*, *uav0000380_00001_s*, *uav0000211_00000_s*, *uav0000094_02070_s*, *uav0000376_00001_s*, *uav0000075_01056_s*, *uav0000286_00001_s*, *uav0000368_00001_s*, *uav0000094_00000_s*, *uav0000381_00001_s*, *uav0000386_00001_s*, *uav0000121_00516_s*, *uav0000298_01242_s*, *uav0000373_00001_s*, *uav0000083_00783_s*, *uav0000154_00099_s*, *uav0000162_00000_s*, *uav0000011_00345_s*, *uav0000096_00345_s*, *and uav0000377_00001_s*. This data set are challenging in 12 aspects, followed by Aspect Ratio Change, Background Clutter, Camera Motion, Fast Motion, Full Occlusion, Illumination Variation, Low Resolution, Out-of-View, Partial Occlusion, Similar Object, Scale Variation and Viewpoint Change.

Evaluation Metrics. According to the official evaluation toolbox, the performance of the proposed tracker, other competitors' trackers and existed trackers are all evaluated by the success and precision scores [38, 41]. Success score reflects the area under the curve (AUC) based on the percentage of successfully tracked frames vs. the bounding box overlap threshold. Precision score reflects whether the distance between the center of the bounding box of tracked frame and the center of bounding box of the ground-truth frame is within 20 pixels. The success score is used to rank the tracking methods.

5.2. Comparison with Existed Trackers

We evaluate the state-of-the-art performance of the proposed method (ATOMFR (RL + InF)) with several existed popular tracking methods which perform well on several natural SOT data sets. These popular methods are MDNet [29], DSiam [20], ECO [6], SiameseFC [2], [28], SRDCF [9], Staple [1], TRACA [5], LCT [26], DSST [11] and KCF [16] and so on. Among these methods, there are DCF based tracking methods, such as KCF, Staple, and RPN based tracking methods, such as DSiam.

As shown in Fig. 8 and Fig. 9, our proposed method (ATOMFR (RL + InF)) ranks the first on all these 60 test sequences no matter on precision score and success score. In terms of precision, our method surpassed the second method of 26.3%, and our method surpassed the sec-

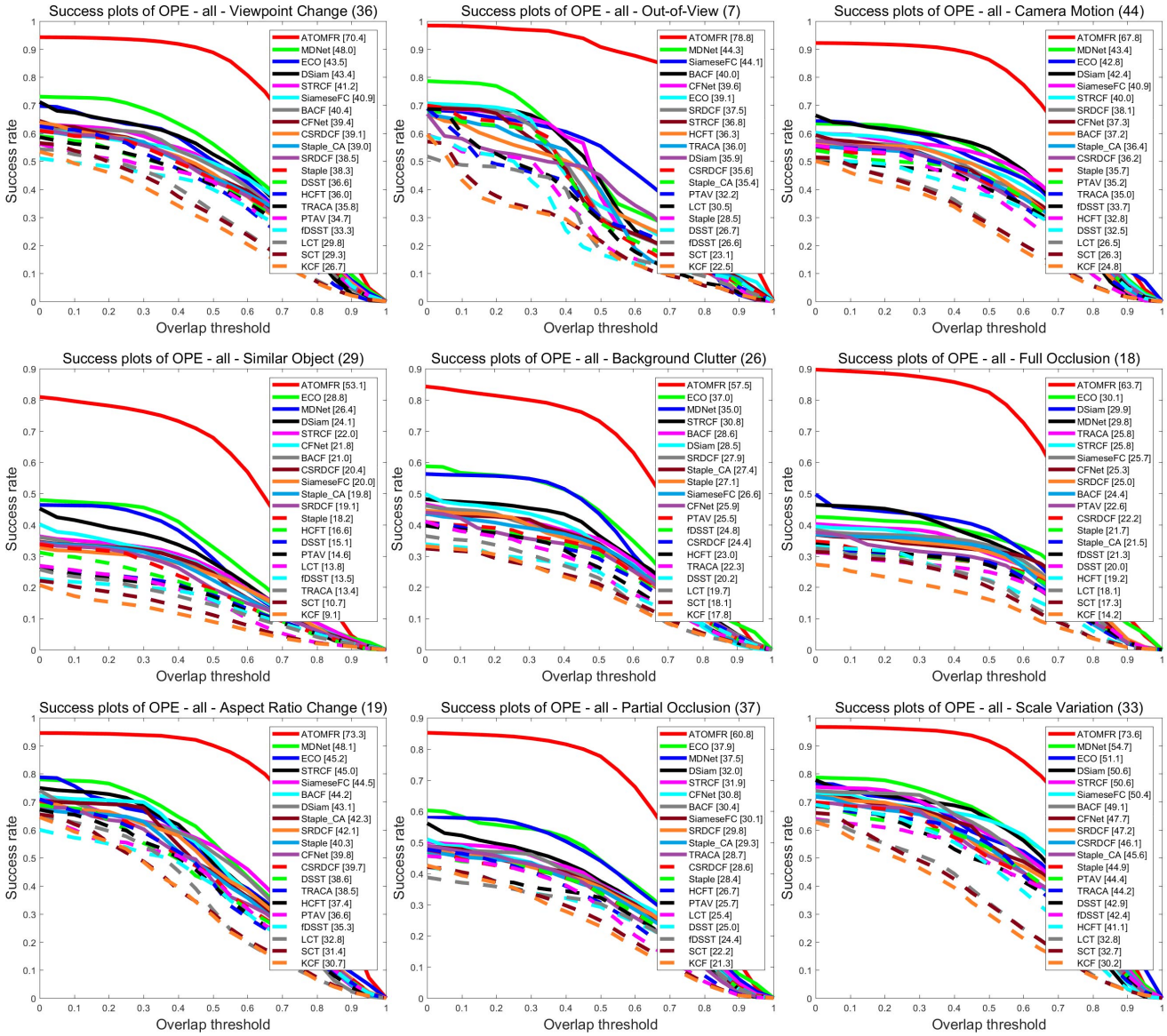

Figure 10. Average performance on VisDrone2019-SOT test set for 9 attributes.

ond ranked method with 19.5% in terms of success rate. This demonstrates the superior performance of our tracking method.

Fig. 10 shows the average performance of ATOMFR (RL + InF) and several existed trackers on VisDrone2019-SOT test set for 9 attributes/aspects (Viewpoint Change, Out-of-View, Camera Motion, Similar Object, Background Clutter, Full Occlusion, Aspect Ratio Change, Partial Occlusion and Scale Variation). From Fig. 10, we can see that ATOMFR (RL + InF) ranks the first on all these 9 attributes, and surpasses the second on attribute Viewpoint Change with 22.4% in terms of success score, on attribute Out-of-View with 34.5% in terms of success score, on attribute Similar Object with 24.3% in terms of success score, on attribute Background Clutter with 20.5% in terms of success score, on attribute Full Occlusion with 33.6% in terms of success score, on attribute Aspect Ratio Change with 15.2% in terms of success score, on attribute Partial Occlusion with 22.9% in terms of success score, and on attribute Scale Variation with 18.9% in terms of success score. There are three main reasons for this performance improvement. First, ATOMFR (RL + InF) uses SE-ResNet-18 as the backbone to train IoU-Net (SE-IoU-Net) which can improve the performance of SE-IoU-Net by recalibrating more useful features and suppressing more useless features. Second, ATOMFR (RL + InF) improve the location

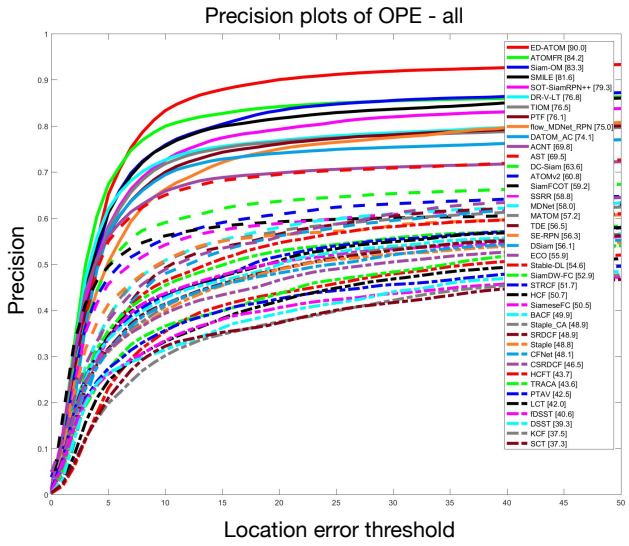

Figure 11. Precision plot on VisDrone2019-SOT test set with other competitors.

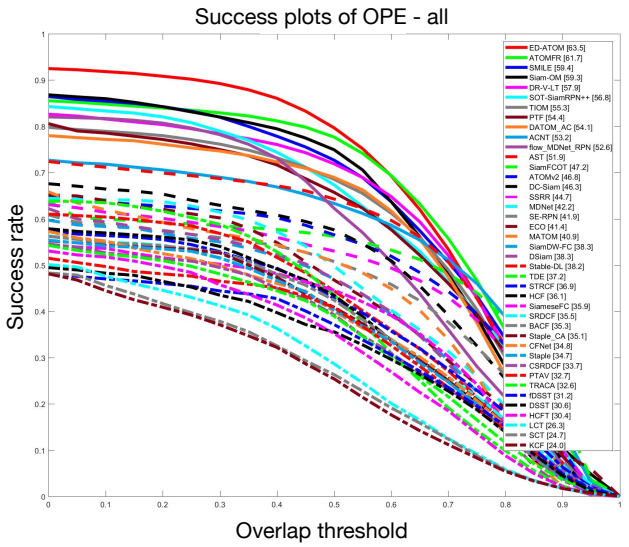

Figure 12. Success plot on VisDrone2019-SOT test set with other competitors.

accuracy by introducing the Relocation Module to relocate the tracked object (usually incorrect located). Last but not least, based on the assumption that the object moves at a constant speed in a short time, the Inference Module introduced in ATOMFR (RL + InF) can infer the position of the object in a complex scene (occulsion). Meanwhile, ATOMFR (RL + InF) uses a DCF based branch (the target classification network) to online discriminate object during tracking process, which gives full use of the discrimination performance of DCF based trackers.

5.3. Comparison with Competitors

In Fig. 11 and Fig. 12, ATOMFR represents our proposed tracker. Here, ATOMFR is the tracker ATOMFR with the Relocation Module (Sec. 4.2) and the Inference Module (Sec. 4.3) (ATOMFR (RL + InF)). It can be seen from Fig. 11 and Fig. 12 that ATOMFR (RL + InF) ranks the second place among all contestants regardless of precision score with 84.7% or success score with 61.7%. This demonstrates the state-of-the-art performance (the top 20%) of our proposed tracker ATOMFR (RL + InF), namely, the collaboration performance of ATOMFR with the Relocation Module and the Inference Module.

5.4. Conclusions and Future Work

In this paper, to improve the performance of the target estimate network (IoU-Net) in ATOM, we introduce the SE-ResNet blocks into the IoU-Net (SE-IoU-Net) and builds the new tracker ATOMFR. ATOMFR enhances the performance of ATOM by the way of feature recalibration (recalibrate more useful features and suppress more useless features). In order to further improve the performance of ATOMFR (scale adaptive ability), we introduce the Relocation Module in ATOMFR and get the tracker ATOMFR (RL). To solve the problem of occlusion in drone based data set (VisDrone2019-SOT), we introduce the Inference Module on the basis of ATOMFR (RL) and obtain the tracker ATOMFR (RL + InF). The experimental results demonstrate the effectiveness of the proposed method (ATOMFR (RL + InF)) on VisDrone2019-SOT test set. By comparing the proposed method with several existing popular tracking methods, the superior performance of the proposed method is verified. By comparing the proposed method with other competitors, the state-of-the-art performance of the proposed method is further demonstrated.

In addition, ATOMFR (RL + InF) is not very good at some specific aspects (such as Illumination Variation), and in the future, we intend to study specific methods/strategies for this challenge.

References

[1] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. Torr. Staple: Complementary Learners for Real-time Tracking. In *CVPR*, 2016.

[2] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. S. Torr. Fully-Convolutional Siamese Networks for Object Tracking. In *ECCV*, 2016.

[3] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui. Visual Object Tracking using Adaptive Correlation Filters. In *CVPR*, 2010.

[4] M. Chao, J. B. Huang, X. Yang, and M. H. Yang. Hierarchical Convolutional Features for Visual Tracking. In *ICCV*, 2016.

[5] J. Choi, H. J. Chang, T. Fischer, S. Yun, K. Lee, J. Jeong, Y. Demiris, and Y. C. Jin. Context-aware Deep Feature Compression for High-speed Visual Tracking. In *CVPR*, 2018.

[6] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg. ECO: Efficient Convolution Operators for Tracking. In *CVPR*, 2017.

[7] M. Danelljan, G. Bhat, F. S. Khan, and M. Felsberg. ATOM: Accurate Tracking by Overlap Maximization. *CoRR*, abs/1811.07628, 2018.

[8] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. Convolutional Features for Correlation Filter Based Visual Tracking. In *ICCVW*, 2015.

[9] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. Learning Spatially Regularized Correlation Filters for Visual Tracking. In *ICCV*, 2015.

[10] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg. Adaptive Decontamination of the Training Set: A Unified Formulation for Discriminative Visual Tracking. In *CVPR*, 2016.

[11] M. Danelljan, G. Haumlger, K. F. Shahbaz, and M. Felsberg. Accurate Scale Estimation for Robust Visual Tracking. In *BMVC*, 2014.

[12] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking. In *ECCV*, 2016.

[13] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *CVPR*, 2016.

[14] Z. Hong, C. Zhe, C. Wang, M. Xue, D. Prokhorov, and D. Tao. MUlti-Store Tracker (MUSTer): a Cognitive Psychology Inspired Approach to Object Tracking. In *CVPR*, 2015.

[15] J. Hu, L. Shen, and G. Sun. Squeeze-and-Excitation Networks. In *CVPR*, 2018.

[16] H. JF, C. R, M. P, and B. J. High-speed Tracking with Kernelized Correlation Filters. *TPAMI*, 2015.

[17] B. Jiang, R. Luo, J. Mao, T. Xiao, and Y. Jiang. Acquisition of Localization Confidence for Accurate Object Detection. In *ECCV*, 2018.

[18] M. Kristan, A. Leonardis, J. Matas, M. Felsberg, R. Pflugfelder, L. C. Zajc, T. Vojir, G. Hager, A. Lukezic, and A. Eldesokey. The Visual Object Tracking VOT2017 Challenge Results. In *ICCVW*, 2017.

[19] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan. SiamRPN++: Evolution of Siamese Visual Tracking with Very Deep Networks. *CoRR*, abs/1812.11703, 2018.

[20] B. Li, J. Yan, W. Wu, Z. Zhu, and X. Hu. High Performance Visual Tracking with Siamese Region Proposal Network. In *CVPR*, 2018.

[21] Y. Li and J. Zhu. A Scale Adaptive Kernel Correlation Filter Tracker with Feature Integration. In *ECCVW*, 2014.

[22] P. Liang, E. Blasch, and H. Ling. Encoding Color Information for Visual Tracking: Algorithms and Benchmark. *IEEE Transactions on Image Processing*, 24(12):5630–5644, 2015.

[23] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan. Discriminative Correlation Filter with Channel and Spatial Reliability. In *CVPR*, 2017.

[24] A. Lukezic, T. Vojir, L. C. Zajc, J. Matas, and M. Kristan. Discriminative Correlation Filter with Channel and Spatial Reliability. In *CVPR*, 2017.

[25] C. Ma, J. B. Huang, X. Yang, and M. H. Yang. Hierarchical Convolutional Features for Visual Tracking. In *ICCV*, 2015.

[26] C. Ma, X. Yang, C. Zhang, and M. H. Yang. Long-term Correlation Tracking. In *CVPR*, 2015.

[27] K. Matej, L. Ales, M. Jırı, F. Michael, P. Roman, and C. Z. Luka. The sixth Visual Object Tracking VOT2018 Challenge Results. In *ECCVW*, 2018.

[28] M. Mueller, N. Smith, and B. Ghanem. Context-aware Correlation Filter Tracking. In *CVPR*, 2017.

[29] H. Nam and B. Han. Learning Multi-domain Convolutional Neural Networks for Visual Tracking. In *CVPR*, 2016.

[30] J. Ning, J. Yang, S. Jiang, L. Zhang, and M. H. Yang. Object Tracking via Dual Linear Structured SVM and Explicit Feature Map. In *CVPR*, 2016.

[31] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards Teal-Time Object Detection with Region Proposal Networks. In *NIPS*, pages 91–99, 2015.

[32] Y. Song, M. Chao, L. Gong, J. Zhang, R. Lau, and M. H. Yang. CREST: Convolutional Residual Learning for Visual Tracking. In *ICCV*, 2017.

[33] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning. In *AAAI*, 2017.

[34] N. Wang and D. Y. Yeung. Real-time Tracking via On-line Boosting. In *ICML*, 2014.

[35] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li. Multi-cue Correlation Filters for Robust Visual Tracking. In *CVPR*, 2018.

[36] Q. Wang, L. Zhang, L. Bertinetto, W. Hu, and P. H. S. Torr. SiamMask: Fast Online Object Tracking and Segmentation. In *CVPR*, 2019.

[37] S. Xie, R. B. Girshick, P. Dollár, Z. Tu, and K. He. Aggregated Residual Transformations for Deep Neural Networks. In *ICCV*, 2017.

[38] W. Yi, J. Lim, and M. H. Yang. Online Object Tracking: A Benchmark. In *CVPR*, 2013.

[39] J. Zhang, S. Ma, and S. Sclaroff. MEEM: Robust Tracking via Multiple Experts Using Entropy Minimization. In *ECCV*, 2014.

[40] T. Zhang, C. Xu, and M. H. Yang. Multi-task Correlation Particle Filter for Robust Object Tracking. In *CVPR*, 2017.

[41] P. Zhu, L. Wen, X. Bian, H. Ling, and Q. Hu. Vision Meets Drones: A Challenge. *CoRR*, abs/1804.07437, 2018.