



WATB: Wild Animal Tracking Benchmark

Fasheng Wang^{1,2} · Ping Cao¹ · Fu Li¹ · Xing Wang¹ · Bing He¹ · Fuming Sun¹

Received: 4 April 2022 / Accepted: 5 December 2022 / Published online: 22 December 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

With the development of computer vision technology, many advanced computer vision methods have been successfully applied to animal detection, tracking, recognition and behavior analysis, which is of great help to ecological protection, biodiversity conservation and environmental protection. As existing datasets applied to target tracking contain various kinds of common objects, but rarely focus on wild animals, this paper proposes the first benchmark, named Wild Animal Tracking Benchmark (WATB), to encourage further progress of research and applications of visual object tracking. WATB contains more than 203,000 frames and 206 video sequences, and covers different kinds of animals from land, sea and sky. The average length of the videos is over 980 frames. Each video is manually labelled with thirteen challenge attributes including illumination variation, rotation, deformation, and so on. In the dataset, all frames are annotated with axis-aligned bounding boxes. To reveal the performance of these existing tracking algorithms and provide baseline results for future research on wild animal tracking, we benchmark a total of 38 state-of-the-art trackers and rank them according to tracking accuracy. Evaluation results demonstrate that the trackers based on deep networks perform much better than other trackers like correlation filters. Another finding on the basis of the evaluation results is that wild animals tracking is still a big challenge in computer vision community. The benchmark WATB and evaluation results are released on the project website <https://w-1995.github.io/>.

Keywords Computer vision · Visual object tracking · Benchmark · Wild animal tracking · Biodiversity conservation

Communicated by Hyun Soo Park.

P. Cao, F. Li, X. Wang, These authors contributed equally to this work.

✉ Fuming Sun
sunfuming@dlnu.edu.cn

Fasheng Wang
wangfasheng@dlnu.edu.cn

Ping Cao
caoping7464@163.com

Fu Li
fuliytu@163.com

Xing Wang
dlnuwangxing@gmail.com

Bing He
hebing_1999@163.com

¹ School of Information and Communication Engineering, Dalian Minzu University, No. 18, Liaohe West Road, Dalian 116600, Liaoning, China

² School of Information Science and Technology, University of Science and Technology of China, No. 96, Jinzhai Road, Hefei 230026, Anhui, China

1 Introduction

The development of machine learning and computer vision makes it possible for ecologists to use computer vision technologies to extract key features from images, videos or other visual data, so as to quickly classify animal species and count individual animals (Tuia et al., 2022; Weinstein, 2018). And these scientists also expect to do further research on animal behavior analysis on the basis of the visual information collected from large datasets (Van der Zande et al., 2021; Norouzzadeh et al., 2018), which consequently facilitates the development of biodiversity conservation (Mathis et al., 2018; Valletta et al., 2017).

Visual object tracking, being one of the key technologies in computer vision, aims to consistently capture interested targets from video sequences (Lu & Wang, 2019). It is a fundamental vision task in many real-world applications including animal recognition and behavior analysis (Lopez-Marcano et al., 2021; Valletta et al., 2017; Risse et al., 2017; Haalck et al., 2020; Fukunaga et al., 2015; Ravoov & Sudarshan, 2020). In the past decades, researchers from the computer vision community have intensively studied visual

object tracking, leaving two main-stream methods, namely correlation filter (CF) based trackers (Liu et al., 2021) and deep learning based trackers (Li et al., 2018). Following the development of tracking methods, tracking benchmarks play an important role in trackers evaluation. It is well known that OTB50 (Wu et al., 2013) is the first benchmark, and has been widely adopted in the literature as a standard tool to evaluate the performance of different trackers. It contains 50 video sequences that cover various kinds of general objects, including animals, vehicles, humans, body parts, toys, etc. In Wu et al. (2015), the authors extend OTB50 to 100 video sequences, which provides a much bigger dataset for trackers evaluation. In the following years, several other benchmark datasets are released, including TC128 (Liang et al., 2015), NFS (Galoogahi et al., 2017), LaSOT (Fan et al., 2019), UAV123 (Mueller et al., 2016), GOT10K (Huang et al., 2021), AMTSet (Wang et al., 2022) and TOTB (Fan et al., 2021). These datasets have promoted the development of visual object tracking.

However, when it comes to wild animals tracking, existing benchmarks are restricted due to the following reasons. First, wild animals cover a wide scope of objects ranging from land animals to marine animals, from insects to mammals. Existing benchmarks do not include enough wild animal sequences to evaluate the performance of a tracker. Second, some existing benchmarks include several video sequences related to animals, such as GOT10K and LaSOT, but most of them are captive animals in zoos which are of less significance than animals captured in the wild. In addition, most of the sequences containing wild animals in those benchmarks are very short which is insufficient for evaluating wild animals tracking. Third, the main challenges in wild animals tracking are different from that of general object tracking. According to our exploration, challenges like camera motion, similar object, deformation and rotation are very common in wild animals tracking. When analyzing group of animals, challenges like full or partial occlusions, out of view and rotation are frequently appeared. For individual animal analysis, rotation and scale variation commonly exist.

In this paper, we build the first benchmark for wild animal tracking (WATB). WATB comprises 206 sequences and covers a wide range of animal categories. The total frame of WATB is over 203,000 while the average sequence length is over 980 frames. All the sequences are labelled with 13 challenge attributes and manually annotated with axis-align bounding boxes. To the best of our knowledge, WATB is the first benchmark dataset specifically used for the research of wild animal tracking. Sample frames from WATB is shown in Fig. 2. We benchmark 38 state-of-the-art trackers for the sake of understanding the performance of them and providing basic results for comparison in future research. Comprehensive analysis of the evaluation results demonstrates that wild

animal tracking remains challenging in computer vision community.

The main contributions of our work are summarized as follows:

- (1) We propose the first benchmark dataset WATB dedicated to the research of wild animal tracking and behavior analysis. WATB consists of 206 sequences with over 203,000 frames, which makes it large enough to support related research. It covers a wide range of wild animal species including molluscs, coelenterates, mammals, birds, reptiles, amphibians, arthropods and fish.
- (2) We evaluate 38 state-of-the-art trackers including deep learning based trackers, such as Siamese network and Transformer, and correlation filter based trackers with hand-crafted and deep features. We give in-depth analysis in order to provide researchers evaluation results for comparison in future research. We rank the evaluated trackers to specify the top three in tracking accuracy, which facilitates us to point out the research prospect or development trends in wild animal tracking according to the latest progress of tracking methods.
- (3) All the sequences, annotations, attributes, evaluation results and toolkit of WATB are released on the website <https://w-1995.github.io/>. Researchers are encouraged to evaluate their work on WATB.

2 Related Works

2.1 Tracking Datasets

Tracking datasets are very important for the development of visual object tracking. There are two standards to categorize the existing benchmark datasets. In Fan et al. (2021), the existing datasets are divided into two types: generic and specific, while in Valmadre et al. (2018) the authors divide existing datasets into long-term and short-term.

Generic datasets usually contain various objects collected from general scenes, such as pedestrians, human body parts, animals, vehicles, sports players, etc. OTB50 (Wu et al., 2013) and OTB100 (Wu et al., 2015) are two standard generic datasets that are widely used to evaluate current trackers in the literature. Both of them contain 11 challenge attributes, and include both color and gray sequences. TC128 (Liang et al., 2015) is composed of 128 color sequences that focus on testing the impact of color information on tracking performance¹. It also provides 11 challenge attributes for each video sequence. NFS (Galoogahi et al., 2017) aims to evaluate trackers on high frame rate videos. Each video sequence is

¹ For the abbreviations, please refer to Table 7 in the supplementary part.

labelled with 9 attributes. Recently, some large-scale datasets are proposed for training deep models for visual object tracking. LaSOT (Fan et al., 2019) provides 1400 sequences with 1120 and 280 for training and testing, respectively. It covers 70 object categories including vehicles, person, pets, live-stock, human body parts, clothes, captive and wild animals, etc. Most of the animal videos are captured at zoo or home and few of them are wild animals. GOT-10K (Huang et al., 2021) provides about 10,000 video sequences while 9,340 for training, 420 for testing and 180 for validation. But its average sequence length is only about 200 frames. OxUvA (Valmadre et al., 2018) mainly focuses on long-term tracking evaluation with 366 video sequences.

Specific dataset is usually constructed to evaluate tracking performance for specific goals. It is worth noting that our WATB is a specific dataset for wild animals tracking. AMTset (Wang et al., 2022) gives special attention to abrupt motion tracking supplying 50 video sequences and a new evaluation metric for performance evaluation. UAV123 (Mueller et al., 2016) focuses on tracking objects captured using unmanned aerial vehicle (UAV). It includes 123 video sequences that cover several object categories including vehicles, person, buildings, etc. In (Fan et al., 2021), the authors proposed the first dataset for transparent object tracking (TOTB). Li et al. (2019) propose the first RGB-thermal object tracking benchmark, while in Xiong et al. (2020), the first dataset is proposed for hyper-spectrum video object tracking. Wang et al. propose a benchmark dataset WAMI-226 for object tracking in the context of wide area motion imagery (WAMI). WAMI-226 focuses on the main challenges induced by low resolution, low contrast and low frame rate. There are also several datasets developed for face tracking. 300-VW (Shen et al., 2015) is composed of 100 videos which is developed for facial landmark tracking. MobiFace (Lin et al., 2019) dataset consists of 80 sequences which is specially developed for face tracking in mobile scenes. In Risse et al. (2017), Risse et al. build a Wildlife Animal Tracking (WAT) dataset which comprises only 12 videos (11 animal videos and 1 table tennis video) covering a small range of wild animals including 1 Mammal video, 1 Bird video and 9 Arthropod videos. WAT pays special attention to small target animals, such as ants, dung beetles, woodlouse, etc., covering common challenges in object tracking including occlusion, illumination variation, similar objects, clutters, and so on. Compared to WAT, our proposed WTAB contains much more animal videos and covers a wide range of animal species (as shown in Fig. 1).

2.2 Tracking Methods

In general, there are two main branches of tracking methods occupying the community, that are correlation filter and deep learning based methods. The former is famous for its high efficiency while maintaining good accuracy and robustness.

The latter is characterized by its high accuracy and robustness under different challenging situations. Here, we briefly review these two methods.

Correlation filter based tracking methods: MOSSE Bolme et al. (2010) is the first CF-based tracking method which brings researchers to a new direction. It shows high tracking accuracy and speed. Following MOSSE, KCF greatly improves its tracking performance by learning kernelized CF on HOG features (Henriques et al., 2014). Staple (Bertinetto et al., 2016) incorporates HOG and color features to improve KCF, while SAMF (Li & Zhu, 2014) pays much attention to scale estimation in CF based trackers. In order to suppress the annoying boundary effects, SRDCF (Danelljan et al., 2015) adopts a spatial weight coefficient to punish the samples that are far away from the object center. BACF (Kiani Galoogahi et al., 2017) tries to enlarge the sampling area and draws real negative samples from background area which greatly improves the quality of samples and suppresses the boundary effect. STRCF (Li et al., 2018) introduces the temporal regularization into SRDCF to further suppress boundary effect. In Dai et al. (2019), ASRCF is proposed to improve SRDCF by introducing adaptive spatial regularization which adapts to object appearance variation. It learns a location filter and a scale filter using HOG and deep features. GFSDCF (Xu et al., 2019, 2021), based on the spatial consistency of multi-channel features, performs group feature selection across spatial and feature channel dimensions, learns correlation filters with spatial-channel group sparsity, and compresses features from spatial and feature channel dimensions. Li et al. (2020) propose to remove cosine window from CF based trackers with spatial regularization and to replace it with a binary and Gaussian shaped mask functions.

Deep learning based tracking methods: The strong power of deep neural network has brought the object tracking community to a new peak. Convolutional Neural Network (CNN) (Nam & Han, 2016; Wang et al., 2015) and Siamese Network (Bertinetto et al., 2016; Li et al., 2018; Guo et al., 2020) are two popular deep tracking frameworks. MDNet (Nam & Han, 2016) adopts large-scale annotated video sets to train a CNN for general feature representation. Wang et al. (2015) explore the representation ability of the features from different CNN layers. The feature maps of different layers are selected and refined to obtain more discriminative features. SiamFC (Bertinetto et al., 2016) is the first Siamese network based tracker which treats tracking as a similarity learning problem. A Siamese network is trained off-line to localize a sample region in a larger searching region. SiamRPN (Li et al., 2018) is composed of a Siamese subnetwork for feature extraction and a region proposal subnetwork for predicting object position and confidence. SiamRPN++ (Li et al., 2019) adopts ResNet50 as the backbone network to extract features. It largely alleviates the damage of trans-

lation invariant brought by padding, and introduces a deep network into the Siamese network, thus bringing significant improvement in accuracy. SiamBAN (Chen et al., 2020) is similar to SiamRPN++. The difference is that the principle of dilated convolution is introduced in SiamBAN, which increases the receptive field and improves the tracking performance. SiamCAR (Guo et al., 2020) is similar to SiamBAN where the difference is the introduction of centerness branch to better localize the object center. SiamAttn (Yu, 2020) uses deformable Siamese attention network and a region refine module to boost tracking performance.

3 Wild Animal Tracking Benchmark (WATB)

3.1 Video Collection

We aim to collect video sequences containing wild animals that are moving naturally. Generally speaking, documentaries about wild animals contain a lot of footage about moving wild animals. Initially, we decided to select sequences from existing general object tracking benchmarks, but very few wild animals are included in these public benchmarks. Thus, we collect a lot of such raw video sequences from the Internet. These sequences are filtered and edited manually leaving those suitable for tracking analysis. Some of the extra frames in the video are cut out, making the video focus more on a specific wild animal of interest. In addition, we also do some clips on some sequences in order to construct certain tracking challenges. For example, a common challenge in object tracking is abrupt motion induced by rapid camera motion or switching which are common in documentaries. We cut some of the frames to make the wild animals move a larger distance between two adjacent frames.

We also aim to cover a wide range of wild animal species including molluscs, coelenterates, mammals, birds, reptiles, amphibians, arthropods and fish, which are from land, sea, forest, grassland and mountain. At the same time, we try to name the selected sequences based on the species of the target animals in the sequences. Since we are not professional zoologists, we use Baidu image recognition² to help us identify the wild animals in the sequence. For this purpose, we select sample frames containing the target animals, and cut down the complete animal region as the input to Baidu image recognition. The results are carefully analyzed to make sure the recognized animals are the same as that from the video sequence. For some animals in crowded scenes, for example, *AcanthopagrusSchlegelii*² and *PomadasyArgenteus*², the recognition results are not accurate enough. We have to try several times and cut different regions of a frame as input to obtain satisfactory

² <http://graph.baidu.com/>.

Table 1 Summary of statistics of the proposed WATB

Item	WATB	WAT
Number of videos	206	11
Total frames	203,836	3,760
Max frames	9061	710
Min frames	135	75
Avg. frames	989	342
Avg. duration	33 seconds	–
Number of att.	13	8
Frame rate	30 fps	–
Animal species	8	3

recognition results. We eventually collected 206 sequences to form the WATB which is sufficient to evaluate a tracker's performance on wild animal tracking. We give the summary of WATB in Table 1 and compare WATB with WAT.

3.2 Animal Categories

Among the 206 sequences in WATB, 29 of them contain sea animals, 102 contain land animals and 75 contain flying animals (shown in Fig. 1a). Most of the sequences in WATB contain birds and mammals, because the raw video sequences are mainly obtained from documentaries that focus much on mammals and birds. There are 27 sequences containing fish and all of them are marine animals. In Fig. 1b, we give the plot of the distribution of the animal species in WATB. It seems that the distribution of the animals in WATB is lack of balance. But it is worth noting that, there are several standards to categorize animals in zoology. We use a rough standard to divide the 206 sequences into several branches. Sequences of mammals can also be divided into several branches, but detailed categorization is out of the scope of our work. As it is difficult to capture wild animal videos by ourselves, existing documentaries are the main source of WATB. We find that wild animals from Africa are the favorite targets in most of the documentaries. In addition, some of the bird sequences are from social media which are taken by amateur photography enthusiasts.

3.3 Annotation

For the annotation step, we follow the standard in Fan et al. (2019) to provide consistent bounding box annotation when labelling most of the sequences. For a given video sequence, the initial target animal is specified in the first frame. When the target animal appears in a frame, the labeler will draw a bounding box tightly including any visible part of the animal. When it is fully occluded or out of view, the labeler will assign an absent label to this frame. However, in order to build

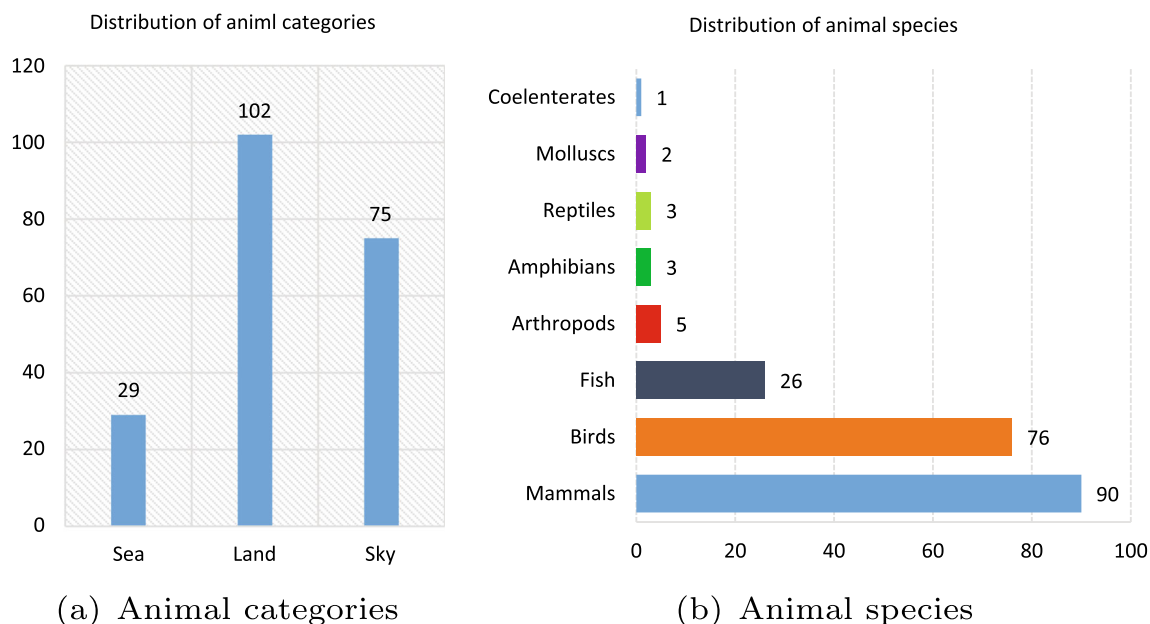


Fig. 1 Distribution of animals in WATB

certain challenge attributes in some sequences, we do not label the whole body of the target animal, because labelling the whole body cannot reflect certain challenge attributes. For example, in Fig. 2, for the sequence *GoldenEagle*, we annotate the head of the bird. For some animals with long tails or legs, take arthropods for example, annotating all the visible parts will include too much background area. Therefore, we only annotate the main body of such animals by omitting the extra legs or tails. For example, in Fig. 2, row 3 column 4, the right leg of the mantis is discarded from the bounding box. Fig. 3 shows the distribution of the bounding box sizes (target animal scale) in WATB. We plot the width (pixels) and height (pixels) of the target animal in the first frame of each sequence. It shows that the widths of the animals approximately fall in the interval [30, 210], while the heights fall in [30, 150].

In order to ensure the annotation quality, we organize a labelling team and an inspection team. Both of them have five members. The labelling team first annotates a sequence and sends the labelled sequence to the inspection team. Then, the inspection members will check all the annotated frames, and any possible inappropriate annotation will be discussed and sent back to the labelling team for refinement. Based on the above principles and quality control measures, we finally get the WATB benchmark dataset. Sample frames selected from WATB are shown in Fig. 2.

3.4 Attributes

In order to enable researchers to do further in-depth analysis of different trackers' performance on wild animal tracking,

we label each video sequence with 13 challenge attributes, that are: (1) illumination variation (IV), (2) out-of-plane rotation (OPR), (3) in-plane rotation (IPR), (4) deformation (DEF), (5) fast motion (FM), (6) scale variation (SV), (7) camera motion (CM), (8) out of view (OV), (9) partial occlusion (POC), (10) full occlusion (FOC), (11) low resolution (LR), (12) similar object (SO), (13) motion blur (MB). The definitions of these attributes are listed in Table 2 which are similar to that in OTB100 and LaSOT. In Fig. 4, the distribution of different challenge attributes in WATB is shown. From this figure, it is clear that the top-ranked common challenges in wild animal tracking are scale variation, out-plane rotation and partial occlusion. The animal behaviors in the collected sequences influence the attributes of the sequences, which can be directly reflected in documentaries. For example, many animals tend to rotate their heads or body out of the image plane. Consequently, the OPR attribute appears in such sequences.

4 Evaluation

4.1 Evaluated Trackers

For the sake of evaluating existing trackers' performance on WATB, we test 38 tracking methods including CF and deep learning based trackers. For CF based trackers, we select the tracking methods with hand-crafted features, such as BACF (Kiani Galoogahi et al., 2017), STRCF (Li et al., 2018), SRDCF (Danelljan et al., 2015), and with deep features, such as ASRCF (Dai et al., 2019), CF2 (Ma et al., 2015), GFS-

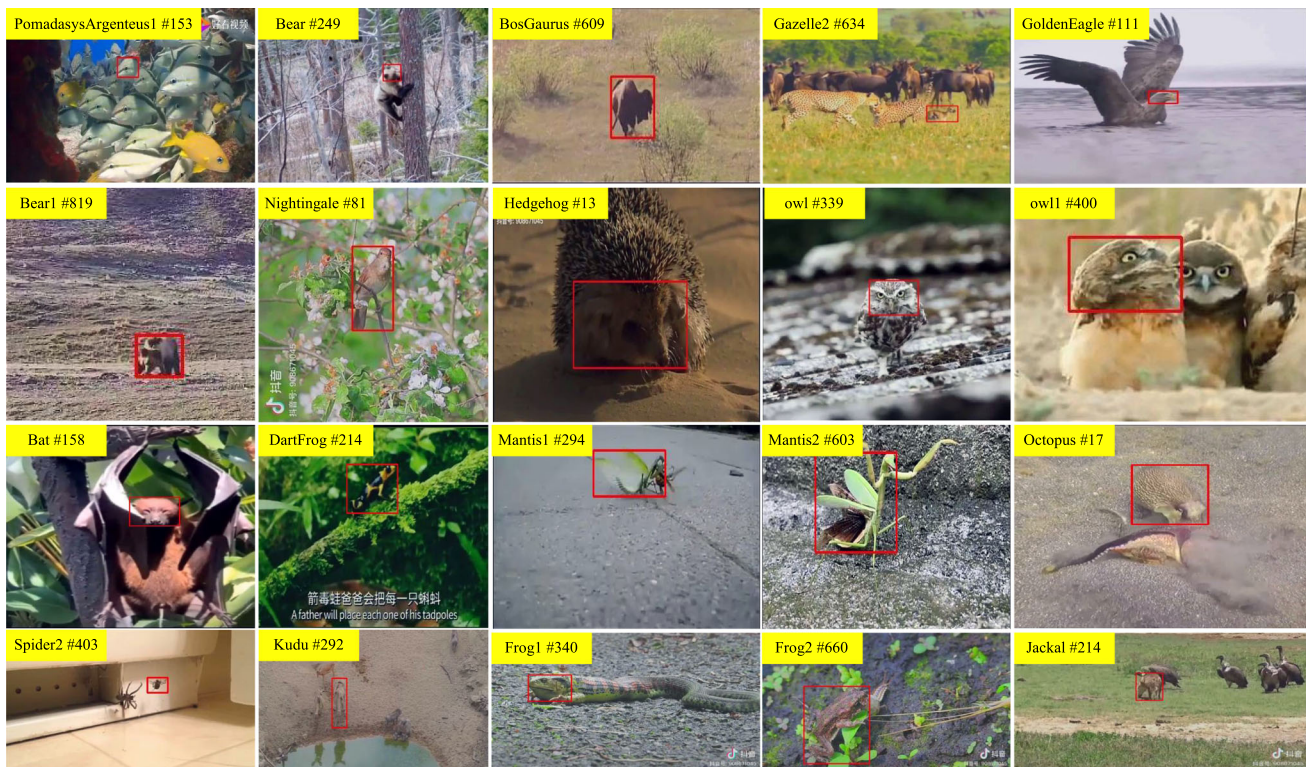


Fig. 2 Sample frames from WATB benchmark

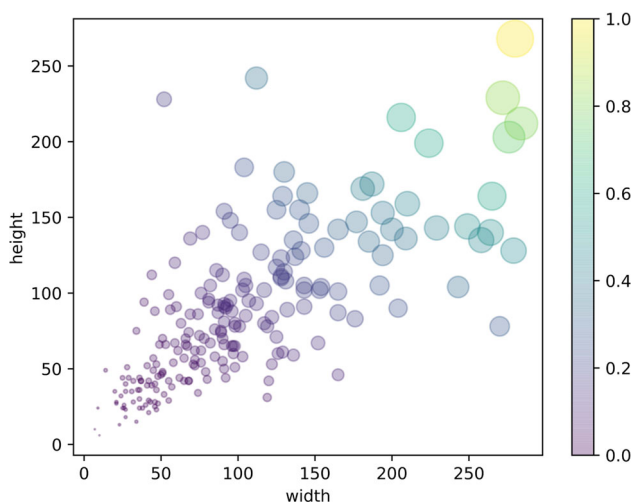


Fig. 3 Distribution of bounding box sizes (object scale) in WATB. The diameter of the bubbles represents the area (width×height) of the corresponding bounding boxes

DCF (Xu et al., 2019). As for deep learning based trackers, we select recent popular Siamese network based trackers, such as SiamBAN (Chen et al., 2020), SiamCAR (Guo et al., 2020), SiamAttn (Yu, 2020), SiamGAT (Guo et al., 2021), STMTrack (Fu et al., 2021), and Transformer trackers, such as Stark (Yan et al., 2021), TransT (Chen et al., 2021), HiFT (Cao et al., 2021). All the tested trackers are listed in Table 3.

We run the codes released by the related authors to implement all the trackers.

4.2 Evaluation Methodology

We use the general one-pass evaluation (OPE) criteria as in Fan et al. (2021), Fan et al. (2019) to compare the trackers using precision measure, normalized precision measure and success measure.

The precision measure is computed as the Euclidean distance d between the estimated object center (x_{est}, y_{est}) and the center of the ground-truth bounding box (x_{gt}, y_{gt}) .

$$d = \sqrt{(x_{est} - x_{gt})^2 + (y_{est} - y_{gt})^2} \quad (1)$$

A distance threshold α_d is first chosen to decide a successful track in each frame (for example, 20 pixels). In a given frame, if $d > \alpha_d$, it is defined as a successful track, which means the tracker successfully captures the target animal. Thereafter, for a tracker, the total successful tracks F_{suc} divide by the total frame number of the animal sequence F_{all} will generate the precision score P which is used to rank different trackers.

$$P = \frac{F_{suc}}{F_{all}} \quad (2)$$

Table 2 Description of the attributes in WATB

No.	Attribute	Description
1	CM	Camera switching, sudden camera motion
2	DEF	The shape of the target is changing during tracking
3	FM	The distance between adjacent frames is larger than the target size.
4	FOC	The object is fully occluded.
5	IPR	The object rotates in the image plane
6	IV	The lighting condition changes during tracking
7	LR	The size of the object area is less than 1000 pixels
8	MB	Blurred object region due to camera shaking or fast moving of object
9	OV	The object is disappeared from the target region
10	OPR	The object rotates out of the image plane
11	POC	The object is partially occluded
12	SO	Similar object appeared in the background
13	SV	The size of the object area is beyond the interval [0.5,2]

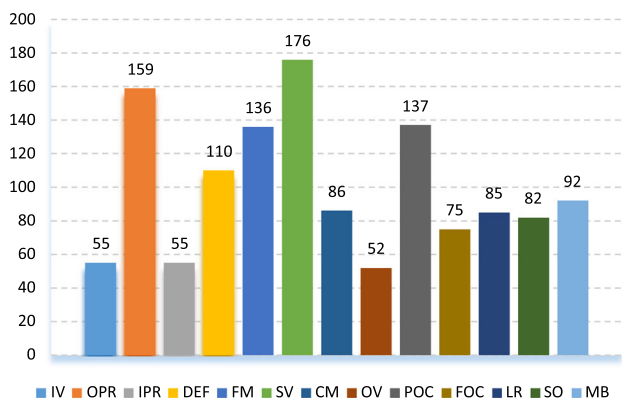


Fig. 4 Sequence numbers of different challenge attributes

As the threshold α_d varies in a given interval [0, 50], a precision curve will be formed. On this plot, the trackers are ranked based on the value of P with distance threshold $\alpha_d = 20$.

The normalized precision measure is adopted because the precision measure is easily influenced by object size. It is computed by normalizing the precision measure over the size of the ground-truth bounding box. The distance between the estimated object center (x_{est}, y_{est}) and the center of the ground-truth bounding box (x_{gt}, y_{gt}) is computed as follow.

$$d_n = \sqrt{\left(\frac{x_{est}}{w} - \frac{x_{gt}}{w}\right)^2 + \left(\frac{y_{est}}{h} - \frac{y_{gt}}{h}\right)^2} \tag{3}$$

where w and h are the width and height of the target animal, respectively. They are defined in ground-truth bounding box. Then, we set a normalized distance threshold α_n to decide a successful track. The definition of the normalized precision score P_n is the same as that of P . To generate a normalized precision curve, α_n varies in a given interval [0, 0.5]. On this plot, the trackers are ranked based on the value of P_n with $\alpha_n = 0.2$.

As for the success measure, it is computed based on the intersection over union (IoU) where the intersection means the overlap area of the estimated region of target animal R_{est} and ground-truth bounding box R_{gt} , and the union means the union of R_{est} and R_{gt} .

$$IoU = \frac{R_{est} \cap R_{gt}}{R_{est} \cup R_{gt}} \tag{4}$$

We also need a threshold α_{IoU} to determine a successful track. Then, the success score S is obtained using the total successful tracks to divide the total number of frames. As the threshold α_{IoU} varies in the given interval [0, 1], a success plot is generated and the trackers on this plot are ranked based on the area under curve (AUC) score.

4.3 Quantitative Results

4.3.1 Overall Performance

We extensively test 38 trackers on WATB. Each tracker is run individually without any modification and re-training. We employ the above-mentioned three measures for all the tested trackers and draw corresponding plots to compare their performance which are shown in Fig. 5.

From Fig. 5, it is clear that Stark obtains the highest precision score (0.580), normalized precision score (0.577) and AUC score (0.512). In the precision plot (Fig. 5a), STMTrack wins the second place (0.556) while TransT wins the second runner-up (0.550). Dimp50 and Dimp18 occupy the fourth and fifth places with precision scores 0.545 and 0.540, respectively. SiamAttn (0.539) and SiamBAN (0.538) are following Dimp50. The other four Siamese network based trackers and ATOM are following the top seven winners. The best CF based tracker is GFSDCF with precision score 0.479 which

Table 3 Evaluated trackers in our experiment

Trackers	Description	Language	Trackers	Description	Language
Stark Yan et al. (2021)	ICCV21	Python	CSK Henriques et al. (2012)	ECCV12	Matlab
STMTrack Fu et al. (2021)	CVPR21	Python	KCF Henriques et al. (2014)	PAMI14	Matlab
TransT Chen et al. (2021)	CVPR21	Python	SAMF Li and Zhu (2014)	ECCVW14	Matlab
Dimp50 Bhat et al. (2019)	ICCV19	Python	DSST Danelljan et al. (2014)	BMVA14	Matlab
Dimp18 Bhat et al. (2019)	ICCV19	Python	fDSST Danelljan et al. (2016)	PAMI16	Matlab
SiamCAR Guo et al. (2020)	CVPR20	Python	CF2 Ma et al. (2015)	ICCV15	Matlab
SiamAttn Yu (2020)	CVPR20	Python	SRDCF Danelljan et al. (2015)	ICCV15	Matlab
ATOM Danelljan et al. (2019)	CVPR19	Python	Staple Bertinetto et al. (2016)	CVPR16	Matlab
SiamRPN++ Li et al. (2019)	CVPR19	Python	CFWCR He et al. (2017)	ICCVW17	Matlab
SiamRPN Li et al. (2018)	CVPR18	Python	BACF Kiani Galoogahi et al. (2017)	ICCV17	Matlab
SiamMask Wang et al. (2019)	CVPR19	Python	ECO Danelljan et al. (2017)	CVPR17	Matlab
SiamGAT Guo et al. (2021)	CVPR21	Python	STRCF Li et al. (2018)	CVPR18	Matlab
SiamBAN Chen et al. (2020)	CVPR20	Python	SKSCF Zuo et al. (2018)	PAMI18	Matlab
HiFT Cao et al. (2021)	ICCV21	Python	DSARCF Feng et al. (2019)	TIP19	Matlab
SiamFC Bertinetto et al. (2016)	ECCV16	Python	LADCF Xu et al. (2019)	TIP19	Matlab
ACSDCF_HC Xu et al. (2021)	IJCV21	Matlab	ASRCF Dai et al. (2019)	CVPR19	Matlab
MRCF Ye et al. (2022)	TIE22	Matlab	GFSDCF Xu et al. (2019)	ICCV19	Matlab
MSCF Zheng et al. (2021)	ICRA21	Matlab	ARCF Huang et al. (2019)	ICCV19	Matlab
AutoTrack Li et al. (2020)	CVPR20	Matlab	DRCF Fu et al. (2020)	TGRS20	Matlab

is better than SiamFC (0.423). The other two deep learning based trackers, SiamGAT and HiFT, do not achieve good precision score because SiamGAT (0.344) focuses much on aspect ratio change while HiFT (0.288) is good at UAV object tracking.

When it comes to the normalized precision plot (Fig. 5b), the top six winners are slightly changed. STMTrack still stands on the second place with score 0.553 followed by TransT (0.547). SiamAttn comes to the fourth place with score 0.539 which is 0.001 higher than Dimp50 (0.538). SiamBAN (0.531) and Dimp18 (0.530) are following Dimp50. The best CF based tracker GFSDCF (0.472) obtains the 13th place and outperforms several deep learning based trackers including SiamFC (0.400) and SiamGAT (0.333). The second best CF based tracker is ASRCF with score (0.384) followed by ARCF(0.348) and CF2 (0.348). As for the success plot, the top five winners do not change much. TransT (0.492) comes to the second place followed by STMTrack (0.486), Dimp50 (0.472) and Dimp18 (0.465). SiamBAN, SiamAttn and ATOM stand on the sixth to eighth places with AUC scores 0.455, 0.455, 0.450, respectively, which outperforms SiamRPN++ (0.428), SiamCAR (0.423), SiamRPN (0.413) and SiamMask (0.395). The best CF based tracker is still GFSDCF (0.373) which outperforms SiamFC (0.331), and the second best CF based tracker is ASRCF (0.324). The rank of SiamGAT (0.322) in the success plot gets a great improvement compared to that in precision plot surpassing all the other CF based trackers.

4.3.2 Attribute-Based Performance

In order to detect the strengths and limitations of existing trackers on wild animal tracking, we provide attribute-based evaluation under the 13 challenge attributes of WATB. In Fig. 6, the plots of the attribute-based performance on WATB are shown which demonstrates Stark is the best tracker on almost all the 13 attributes except similar object and illumination variation. STMTrack wins the champion on similar object attribute.

The most common challenge in WATB is scale variation, and Stark gets the best AUC score (0.504) which outperforms the second best TransT (0.478) by 2.6%. STMTrack, Dimp50 and Dimp18 stand on the following three places. The best two CF based trackers are GFSDCF and ASRCF with AUC scores 0.333 and 0.280, respectively. Out-of-plane rotation is the second most common challenge in wild animal tracking. In this attribute, the top five trackers are almost the same as those in scale variation with TransT (0.470) and STMTrack (0.471) exchanging their places. GFSDCF, ASRCF, CFWCR and CF2 are the best four CF based trackers.

When it comes to partial occlusion, the top five trackers are the same as those in scale variation, but the AUC scores show severe decrease, which demonstrates that partial occlusion remains a big challenge in wild animal tracking. It is worth noting that GFSDCF (0.318) and ASRCF (0.275) are the best two CF based trackers. ASRCF outperforms SiamFC (0.270) and HiFT (0.244) by 0.5% and 3.1%, respectively. On the fast

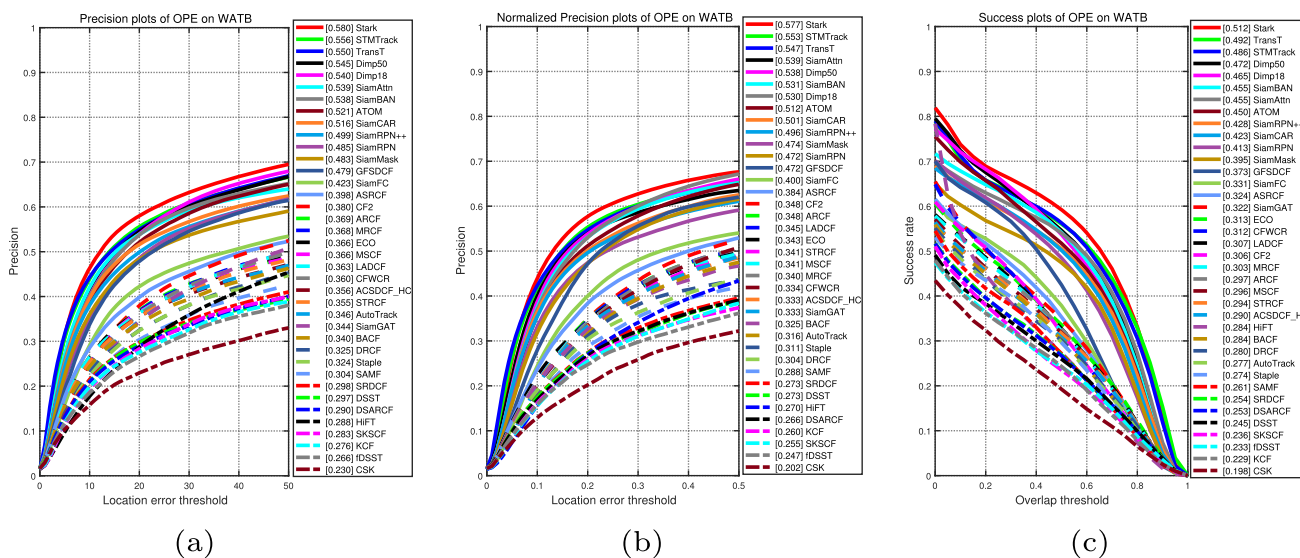


Fig. 5 Overall performance of the trackers on WATB using precision, normalized precision and success measures

motion attribute, the overall AUC scores also show severe decrease compared to scale variation, which means that fast motion is another big challenge in wild animal tracking. Fast motion is commonly induced by camera switching, sudden dynamic change and fast moving of animals. Stark shows the best performance with AUC score 0.465. It outperforms the second best TransT (0.423) by a large margin 4.2%. Dim50 is the second runner up with an AUC score 0.401 which is 0.7% higher than STMTrack and 2.2% lower than TransT (0.423). The AUC score of the best CF based tracker GFSDCF is only 0.287 which is 17.8% lower than Stark. On the out of view attribute, we can draw similar conclusions.

On the full occlusion attribute, the AUC scores of the trackers decrease severely. The best tracker Stark only gets 0.385, which tells us that full occlusion is very challenging in wild animal tracking. We can get the conclusion from the literatures that full occlusion is also challenging in general object tracking. For similar object (background clutters), STMTrack stands on the first place with AUC score 0.480 followed by TransT (0.478) and SiamAttn (0.470), while Stark (0.469) falls to the fourth place. STMTrack wins the second place on illumination variation attribute with AUC score (0.448) which is 1.3% lower than Stark (0.461), and TransT (0.445) is the second runner up exceeding Dimp18 and Dimp50. We can see that, all the deep learning based trackers exceed CF based trackers except the best two trackers, i.e. GFSDCF (0.328) and ASRCF (0.288).

The ranking of the trackers in Fig. 6 clearly tells us that deep learning based trackers are more efficient in dealing with these challenge attributes in wild animal tracking, while CF based trackers, with either deep features or hand-crafted features, are less capable of handling these challenges.

The normalized precision scores of the evaluated trackers over 13 attributes are given in Table 4. The results of deep learning based trackers and CF based trackers are separated and listed in different parts of the table. For deep learning based trackers, the best three results are shown in red, blue and green, respectively, while the best three CF based trackers are shown in italic style using similar colors to deep learning based trackers. The scores show severe drop on FOC attributes, which tells us that FOC is very challenging in wild animal tracking. Similar conclusion can also be drawn from Fig. 6. Stark gets the highest score on almost all of the attributes except SO and IV, while GFSDCF achieves the best on all of the attributes among CF based trackers. For CF based trackers, the top ranked ones, including GFSDCF, ASRCF, CF2 and ECO, use deep features and hand-crafted features together to enhance the tracking performance. The scores of such trackers are better than those CF based trackers with hand-crafted features. ASRCF wins the runner up on most of the attributes except CM. Most of the CF based trackers with hand-crafted features cannot obtain satisfactory results, which demonstrates that deep features have better representation ability in wild animal tracking. We will give further analysis in Sect. 4.4.

4.3.3 Species-Based Performance

As shown in Fig. 1b, there are 8 animal species in WATB. In this section, we conduct species-based evaluation on WATB. The success plots of the evaluation results are shown in Fig. 7. We do not give the success plots of Amphibians, Reptiles, Molluscs and Coelenterates due to the fact that the sequences of these animal species are very few (less than 4), but all of them are released on our project website. Therefore, we only

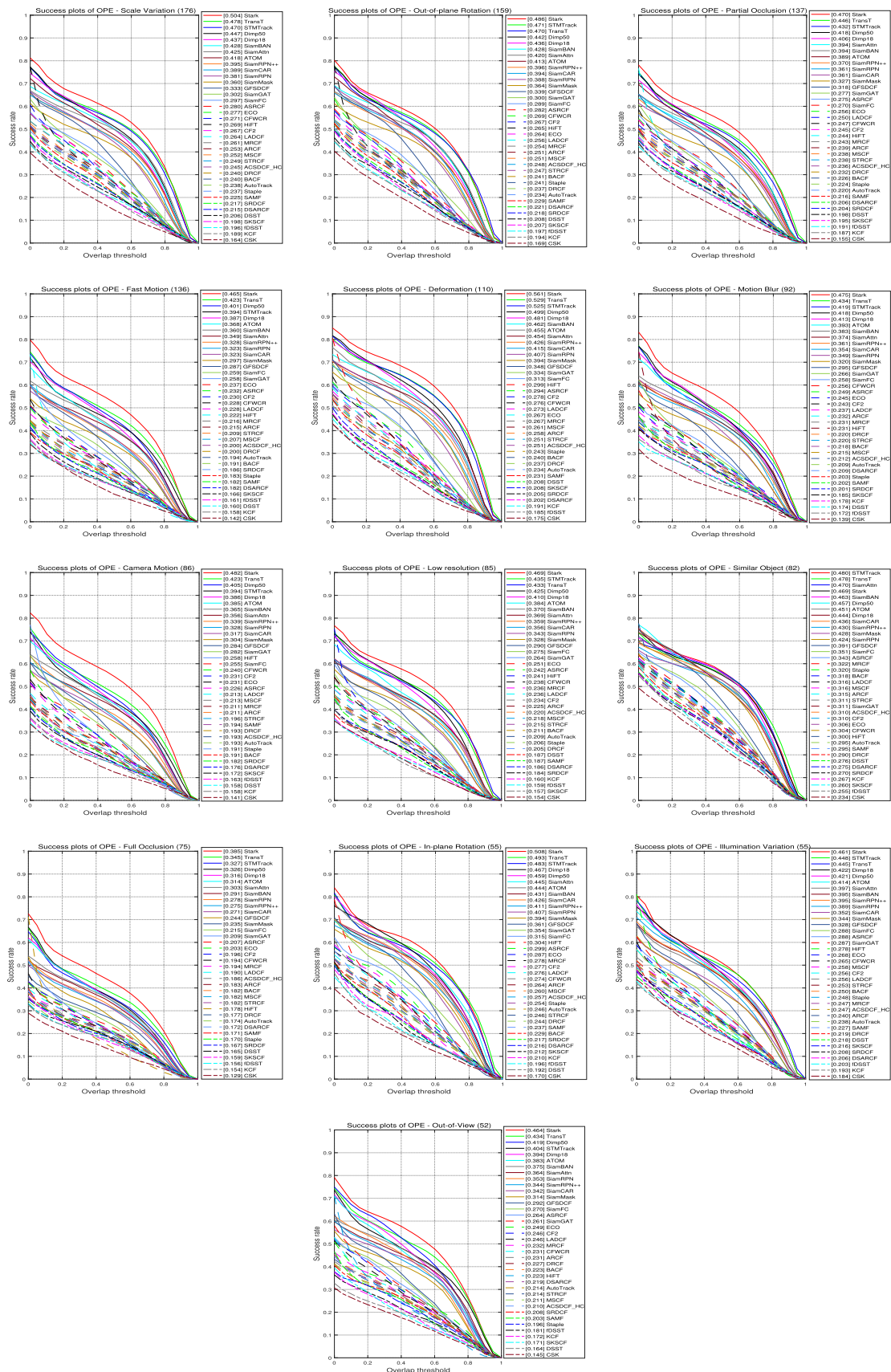


Fig. 6 Attribute-based performance of the trackers on WATB using success measure

Table 4 Normalized precision score of different trackers over 13 challenge attributes (Color table online)

Trackers	SV	OPR	POC	FM	DEF	MB	CM	LR	SO	FOC	IPR	IV	OV
Stark	0.567	0.544	0.531	0.510	0.635	0.517	0.524	0.528	0.537	0.418	0.557	0.503	0.524
STMTrack	0.535	0.535	0.488	0.433	0.598	0.451	0.428	0.496	0.559	0.346	0.538	0.510	0.456
TransT	0.532	0.522	0.494	0.460	0.590	0.459	0.453	0.476	0.548	0.371	0.538	0.483	0.484
Dimp50	0.507	0.492	0.471	0.444	0.572	0.445	0.437	0.480	0.527	0.349	0.506	0.469	0.469
SiamAttn	0.501	0.495	0.464	0.402	0.536	0.426	0.409	0.422	0.566	0.336	0.516	0.464	0.423
SiamBAN	0.496	0.496	0.456	0.403	0.534	0.432	0.402	0.421	0.551	0.322	0.488	0.446	0.439
Diampl8	0.495	0.487	0.463	0.427	0.546	0.450	0.414	0.466	0.510	0.347	0.534	0.459	0.449
ATOM	0.473	0.457	0.433	0.408	0.516	0.425	0.427	0.427	0.520	0.333	0.505	0.467	0.426
SiamRPN++	0.457	0.457	0.428	0.363	0.493	0.400	0.373	0.407	0.512	0.305	0.471	0.452	0.394
SiamCAR	0.455	0.466	0.428	0.371	0.486	0.411	0.356	0.416	0.534	0.311	0.492	0.409	0.419
SiamRPN	0.435	0.440	0.417	0.356	0.466	0.380	0.355	0.382	0.497	0.310	0.451	0.432	0.404
SiamMask	0.430	0.436	0.395	0.341	0.470	0.365	0.344	0.393	0.532	0.277	0.458	0.397	0.376
SiamFC	0.359	0.339	0.326	0.309	0.376	0.297	0.288	0.341	0.430	0.241	0.390	0.350	0.326
SiamGAT	0.312	0.308	0.283	0.252	0.350	0.245	0.279	0.259	0.330	0.197	0.372	0.298	0.255
HiFT	0.251	0.246	0.224	0.195	0.278	0.193	0.232	0.209	0.293	0.142	0.305	0.265	0.193
GFSDCF	0.417	0.417	0.395	0.356	0.427	0.357	0.342	0.382	0.494	0.299	0.454	0.424	0.357
ASRCF	0.329	0.326	0.318	0.258	0.336	0.275	0.243	0.303	0.431	0.236	0.364	0.343	0.307
CF2	0.304	0.295	0.275	0.257	0.310	0.267	0.249	0.265	0.372	0.216	0.317	0.277	0.288
ARCF	0.296	0.286	0.274	0.249	0.296	0.270	0.238	0.273	0.376	0.210	0.321	0.268	0.291
ECO	0.295	0.270	0.268	0.257	0.274	0.249	0.233	0.277	0.336	0.218	0.322	0.281	0.280
LADCF	0.292	0.273	0.271	0.247	0.295	0.259	0.228	0.271	0.361	0.205	0.326	0.282	0.281
STRCF	0.286	0.275	0.268	0.240	0.284	0.240	0.220	0.255	0.369	0.213	0.288	0.302	0.240
MRCF	0.286	0.272	0.271	0.237	0.287	0.257	0.226	0.267	0.367	0.211	0.310	0.269	0.277
MSCF	0.285	0.279	0.271	0.234	0.290	0.248	0.236	0.254	0.373	0.207	0.297	0.294	0.264
CFWCR	0.281	0.268	0.253	0.239	0.274	0.251	0.236	0.268	0.342	0.206	0.277	0.298	0.249
ACSDCF_HC	0.275	0.272	0.256	0.223	0.279	0.226	0.216	0.257	0.381	0.210	0.310	0.285	0.228
AutoTrack	0.274	0.264	0.252	0.224	0.266	0.240	0.218	0.247	0.349	0.196	0.287	0.258	0.271
BACF	0.272	0.270	0.254	0.210	0.259	0.243	0.194	0.250	0.375	0.202	0.248	0.285	0.264
Staple	0.265	0.271	0.252	0.207	0.267	0.228	0.217	0.237	0.368	0.184	0.304	0.283	0.247
DRCF	0.257	0.247	0.255	0.209	0.247	0.226	0.190	0.230	0.318	0.187	0.254	0.211	0.254
SAMF	0.244	0.242	0.232	0.186	0.232	0.216	0.198	0.200	0.332	0.176	0.252	0.236	0.245
SRDCF	0.229	0.220	0.215	0.201	0.210	0.204	0.197	0.206	0.295	0.177	0.215	0.231	0.231
DSST	0.224	0.225	0.213	0.179	0.225	0.182	0.172	0.213	0.317	0.180	0.213	0.226	0.193
DSARCF	0.222	0.219	0.214	0.190	0.201	0.204	0.182	0.196	0.293	0.174	0.203	0.214	0.230
KCF	0.212	0.209	0.206	0.180	0.208	0.180	0.173	0.187	0.305	0.171	0.244	0.220	0.197
SKSCF	0.209	0.215	0.206	0.176	0.207	0.184	0.188	0.171	0.285	0.167	0.217	0.224	0.196
fDSST	0.205	0.198	0.203	0.168	0.184	0.169	0.170	0.173	0.271	0.168	0.206	0.225	0.202
CSK	0.164	0.166	0.162	0.137	0.157	0.135	0.136	0.171	0.251	0.133	0.179	0.187	0.172

discuss the results of the other four species, i.e. Birds, Fish, Mammals and Arthropods.

For Birds sequences, Stark (0.487), TransT (0.476) and Dimp50 (0.474) occupy the top three places all of which are transformer-based trackers. The AUC scores of the following 5 deep learning based trackers, SiamBAN, Dimp18, ATOM, SiamAttn, and STMTrack, are all higher than 0.450. GFS-DCF (0.411), ASRCF (0.352) and CFWCR (0.348) are the best three CF-based trackers on these sequences. The other trackers following GFSDCF get lower AUC scores than 0.4. When it comes to the Fish sequences, STMTrack (0.550) wins the championship, following TransT (0.533) and SiamAttn (0.505). The overall AUC scores of Fish sequences are obviously much higher than Birds sequences, which tells us that Birds sequences contain much more challenge attributes. For Mammals sequences, Stark (0.525) returns to the first place, leading the runner-up STMTrack with a large margin (3.5%). The AUC scores of the trackers behind SiamRPN (0.401) are all lower than 0.400 including all the CF based trackers. There are only 5 Arthropods sequences in WATB which is much fewer than the aforementioned 3 species. Stark undoubtedly occupies the first place leading the runner-up TransT a very large margin (16.7%). This success plot also shows that all the trackers except Stark and TransT have

their AUC values below 0.4, which demonstrates that Arthropods tracking is very challenging in wild animal tracking, but the advanced Transformer-based trackers have shown a very promising potential to overcome this problem.

In Table 5, the normalized precision scores of the trackers over 8 animal species in WATB are shown. We mark the best three scores using the same appearance as that in Table 4. We do not mark the results of two species, Molluscs and Coelenterates, considering that there are very few sequences of these two species. For the marked 6 animal species, the best methods are all Transformer-based trackers, which again demonstrates their great potential in tracking different animal species. For Arthropods, Stark gets the best score (0.637), but the scores of the other 37 trackers show severe drop compared to Stark, especially the CF based trackers. It is very difficult for those CF based trackers to accurately track the target Arthropod animals. This result also verifies the challenge attribute of tracking Arthropod animals and the potential of Transformers in conquering these challenges. On the other 5 animal species, GFSDCF shows moderate performance compared to the corresponding best three deep learning based trackers, and leads the other CF-based trackers by a large margin which can be attributed to the use of deep features as in Sect. 4.3.2.

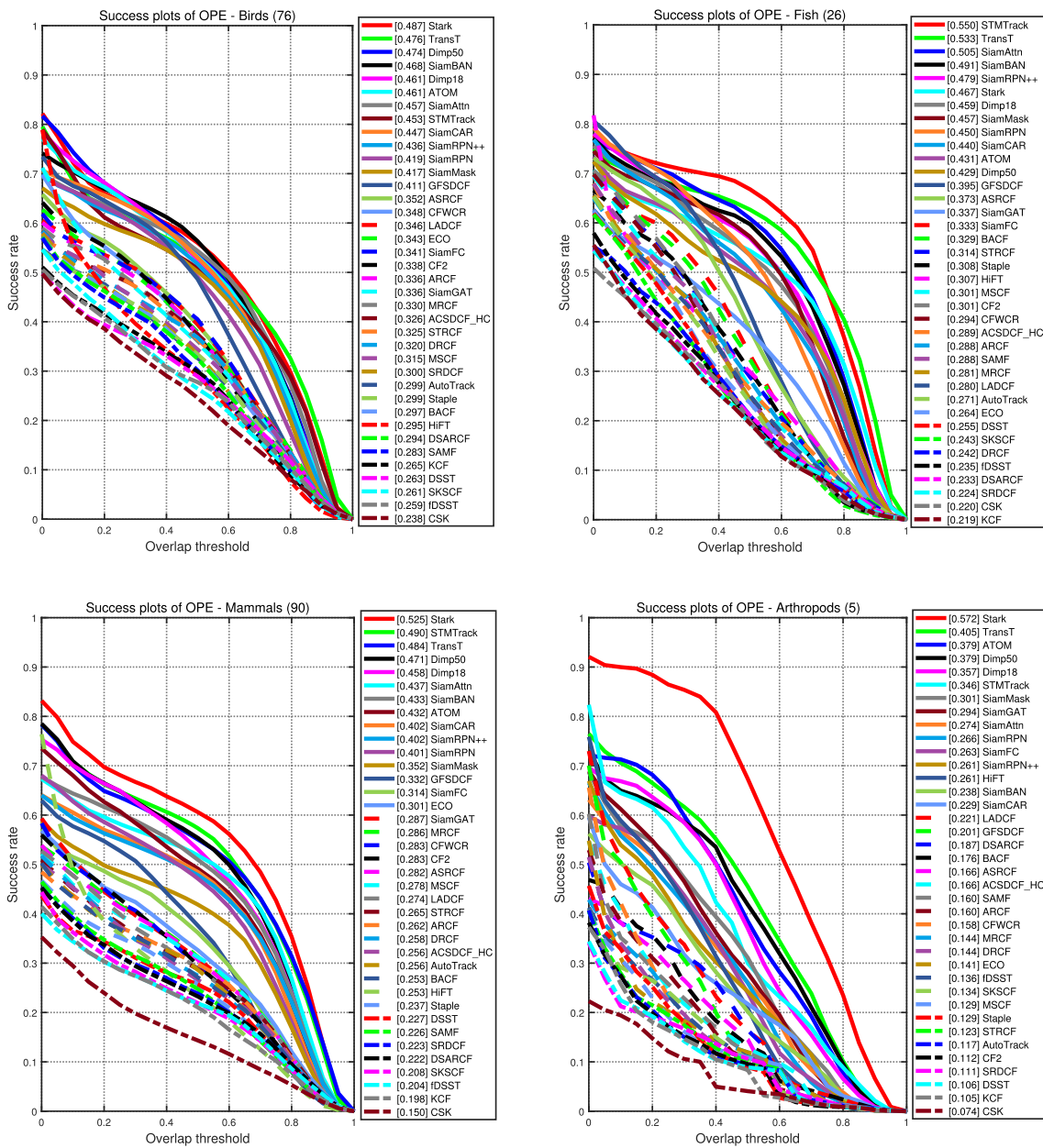


Fig. 7 Species-based performance of the trackers on WATB using success measure

4.4 Deep and Hand-Crafted Features in WATB

As we have claimed in the previous section, compared to CF based trackers with deep features, those with hand-crafted features cannot deal well with the challenges in wild animal tracking. In this section, we do ablation study using GFSDDCF tracker (Xu et al., 2019).

As the baseline, GFSDDCF uses both deep features from ResNet50, and hand-crafted features including HOG and CN. HOG feature is wildly adopted in CF based trackers. CN feature is proven to be better than any other color features in object tracking. The results are shown in Fig. 8 and Table 6,

which tells us that using single hand-crafted feature, HOG or CN, cannot achieve good tracking performance. But when they work together, the tracking performance is boosted with a large margin. Consequently, the three measures increase remarkably. When deep features from ResNet50 are incorporated, the precision score increases by 10.7%, normalized precision score 12.9% and AUC score 7.4%. It clearly demonstrates that using deep features can greatly improve the tracking performance on WATB. Compared to single hand-crafted feature, incorporating multiple hand-crafted features can also boost the tracking performance to some extent.

Table 5 Normalized precision score of different trackers over 8 animal species in WATB (Color table online)

Trackers	Birds	Fish	Arthropods	Amphibians	Mammals	Reptiles	Molluscs	Coelenterates
Stark	0.544	0.546	0.637	0.772	0.589	0.930	0.609	0.778
STMTrack	0.495	0.675	0.409	0.789	0.559	0.667	0.802	0.374
TransT	0.523	0.633	0.426	0.857	0.533	0.690	0.707	0.423
Dimp50	0.529	0.504	0.525	0.730	0.536	0.722	0.691	0.941
SiamAttn	0.541	0.626	0.270	0.697	0.511	0.572	0.813	0.780
SiamBAN	0.544	0.602	0.211	0.807	0.498	0.534	0.802	0.883
Diamp18	0.521	0.544	0.474	0.754	0.515	0.735	0.702	0.842
ATOM	0.524	0.520	0.458	0.664	0.476	0.828	0.736	0.995
SiamRPN++	0.496	0.582	0.236	0.624	0.471	0.529	0.738	0.875
SiamCAR	0.528	0.526	0.225	0.624	0.480	0.423	0.695	0.575
SiamRPN	0.476	0.550	0.221	0.429	0.454	0.496	0.753	0.607
SiamMask	0.494	0.600	0.292	0.492	0.418	0.553	0.740	0.842
SiamFC	0.414	0.410	0.251	0.564	0.374	0.297	0.573	0.791
SiamGAT	0.333	0.403	0.262	0.760	0.289	0.452	0.679	0.465
HiFT	0.282	0.323	0.186	0.686	0.226	0.403	0.490	0.322
GFSDCF	0.520	0.541	0.210	0.629	0.406	0.733	0.515	0.789
ASRCF	0.418	0.456	0.188	0.423	0.334	0.344	0.666	0.791
CF2	0.386	0.332	0.141	0.413	0.324	0.329	0.470	0.737
ARCF	0.394	0.324	0.146	0.463	0.314	0.364	0.587	0.790
ECO	0.378	0.273	0.153	0.523	0.333	0.371	0.338	0.791
LADCF	0.405	0.291	0.227	0.401	0.303	0.493	0.412	0.791
STRCF	0.378	0.347	0.126	0.385	0.312	0.423	0.319	0.776
MRCF	0.374	0.292	0.128	0.392	0.326	0.334	0.588	0.740
MSCF	0.363	0.318	0.135	0.428	0.328	0.298	0.587	0.784
CFWCR	0.379	0.320	0.122	0.450	0.301	0.416	0.310	0.791
ACSDCF_HC	0.378	0.323	0.158	0.418	0.299	0.377	0.303	0.791
AutoTrack	0.343	0.280	0.116	0.389	0.301	0.369	0.587	0.628
BACF	0.344	0.383	0.168	0.394	0.290	0.273	0.586	0.791
Staple	0.351	0.332	0.067	0.418	0.271	0.338	0.585	0.607
DRCF	0.352	0.266	0.134	0.392	0.275	0.264	0.371	0.767
SAMF	0.307	0.298	0.158	0.387	0.257	0.321	0.679	0.791
SRDCF	0.329	0.241	0.082	0.387	0.236	0.312	0.304	0.741
DSST	0.297	0.250	0.131	0.383	0.262	0.245	0.303	0.408
DSARCF	0.311	0.243	0.142	0.394	0.231	0.221	0.312	0.781
KCF	0.305	0.249	0.082	0.385	0.224	0.242	0.327	0.777
SKSCF	0.271	0.263	0.134	0.386	0.230	0.233	0.536	0.791
fDSST	0.277	0.235	0.093	0.384	0.224	0.235	0.230	0.790
CSK	0.235	0.257	0.057	0.382	0.152	0.225	0.344	0.503

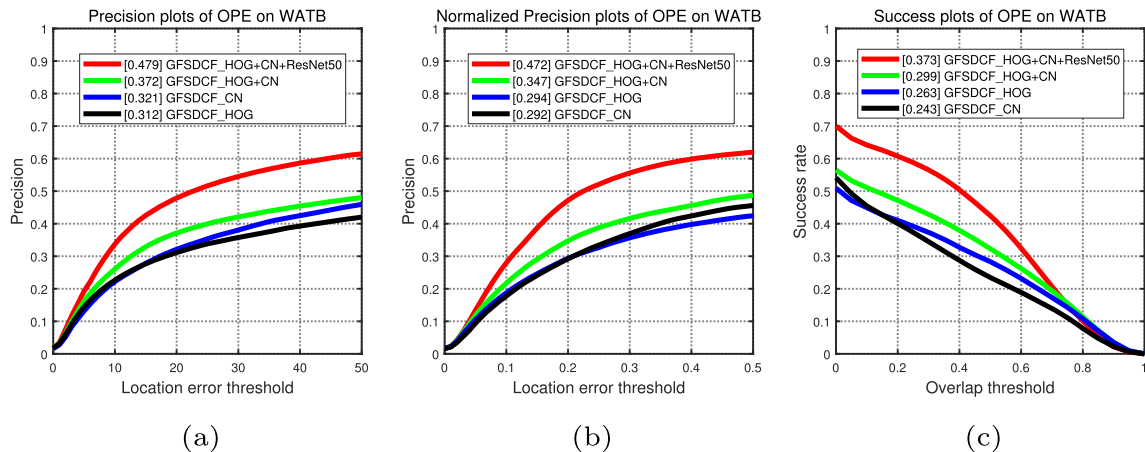


Fig. 8 Plots of GFSDCF on WATB using hand-crafted and deep features

Table 6 Ablation study of deep and hand-crafted features in WATB for GFSDCF

Trackers	Precision	Normalized Precision	Success AUC
GFSDCF_HOG	0.312	0.294	0.263
GFSDCF_CN	0.321	0.292	0.243
GFSDCF_HOG+CN	0.372	0.347	0.299
GFSDCF_HOG+CN+ResNet50	0.479	0.472	0.373

4.5 Qualitative Evaluation

For qualitatively analyzing the existing trackers, we visualize the tracking results of six typical trackers, Stark, TransT, SiamBAN, GFSDCF, ASRCF, STRCF, which covers transformer trackers (Stark and TransT), Siamese network based tracker (SiamBAN), CF based trackers with deep and hand-crafted features (GFSDCF and ASRCF), and CF based trackers with hand-crafted features (STRCF). We select six typical animal sequences, *BabyLion*, *Boar*, *Clownfish4*, *Crane*, *Gazelle1*, *Lynxlynx*. Each of the sequences contains several challenge attributes and different animals. The results are shown in Fig. 9.

For the *BabyLion* sequence, the target animal undergoes frequent rotation, partial occlusion, and similar targets are appeared to disturb tracking. When severe rotations appeared (#0243), Stark, TransT and SiamBAN can accurately capture its position and scale, while the CF based trackers cannot adapt to the scale changes. After long-period partial occlusions (#0988), Stark first recovers from failure and captures the target animal accurately, followed by TransT (#1137), but the other trackers drifts to other similar animals. The *Boar* sequence covers several challenge attributes including motion blur, abrupt motion, out of view, scale variation and out-of-plane rotation. The results show that CF based trackers can not well adapt to scale variations induced by frequent rotations, while the three deep learning based trackers can deal well with these challenges. The *Clownfish4* and *Crane* sequences show some failure cases of the tracker Stark when the target animals frequently interact with similar animals accompanying frequent occlusions, scale variations and deformations. However, Stark can successfully recover from such failures and show better performance in handling these accompanied challenges. In frame #1196 of *Crane* sequence, all the trackers are distracted from the target animal. Similar conclusion can be made on the *Gazelle1* sequence. The *Lynxlynx* sequence is very challenging since it contains severe deformation, scale variation and motion blur. The target animal also does sudden jump to change its direction and shape. We can see that GFSDCF and STRCF drift after the first jump (#0149). After the third (#0179) and fourth jump (#0211), Stark can still stably track the animal. The *Nightingale* sequence contains the severe abrupt motion challenge induced by camera switching. The target animal undergoes

large-scale position change. Given the consecutive frames (#0194–#0197) where the target bird changes its position suddenly with large difference, all the trackers fail to capture it. But Stark quickly recovers from failure and re-capture the bird. TransT recovers at frame #224.

From the above qualitative analysis, we can draw the conclusion that deep learning based trackers can deal with various challenges in wild animal tracking better than CF based trackers. Some of the challenges, for example, abrupt motions induced by camera switching, frequent similar animal interactions, or frequent deformation, scale variation and rotation, cannot be well handled by the CF based trackers. There are some failure cases for the best deep learning based trackers (Stark and TransT), which tells us that wild animal tracking is far from well resolved.

5 Discussion

The main contribution of this paper is to provide the community a benchmark for evaluating trackers and boost the research of wild animal tracking. We have successfully constructed the WATB and released all the resources on the project website. WATB consists of 206 video sequences covering different kinds of animals that are usually appeared in documentaries. According to zoologists, there are about 1.5 million known animals on earth. Therefore, we need to extend WATB to cover more animal species. But finding all of the wild animals in videos seems impossible. We will supplement absent animal species to WATB in our following work. In addition, we do not provide training set in WATB for existing deep learning trackers because the 206 sequences in WATB are very few compared to existing large scale training set. According to our exploration, the existing training sets for training object tracking models are very large, for example, the training set of GOT-10K contains more than 9,300 sequences, LaSOT has more than 1,100 training sequences, and TrackingNet has more than 30,000 training sequences. Most of the existing deep learning based tracking models are trained using the above-mentioned training sets. We will consider to build a training set for wild animal tracking in our future work.

According to our evaluation results, the strongest tracker in WATB is Stark (Yan et al., 2021). But its scores of the nor-

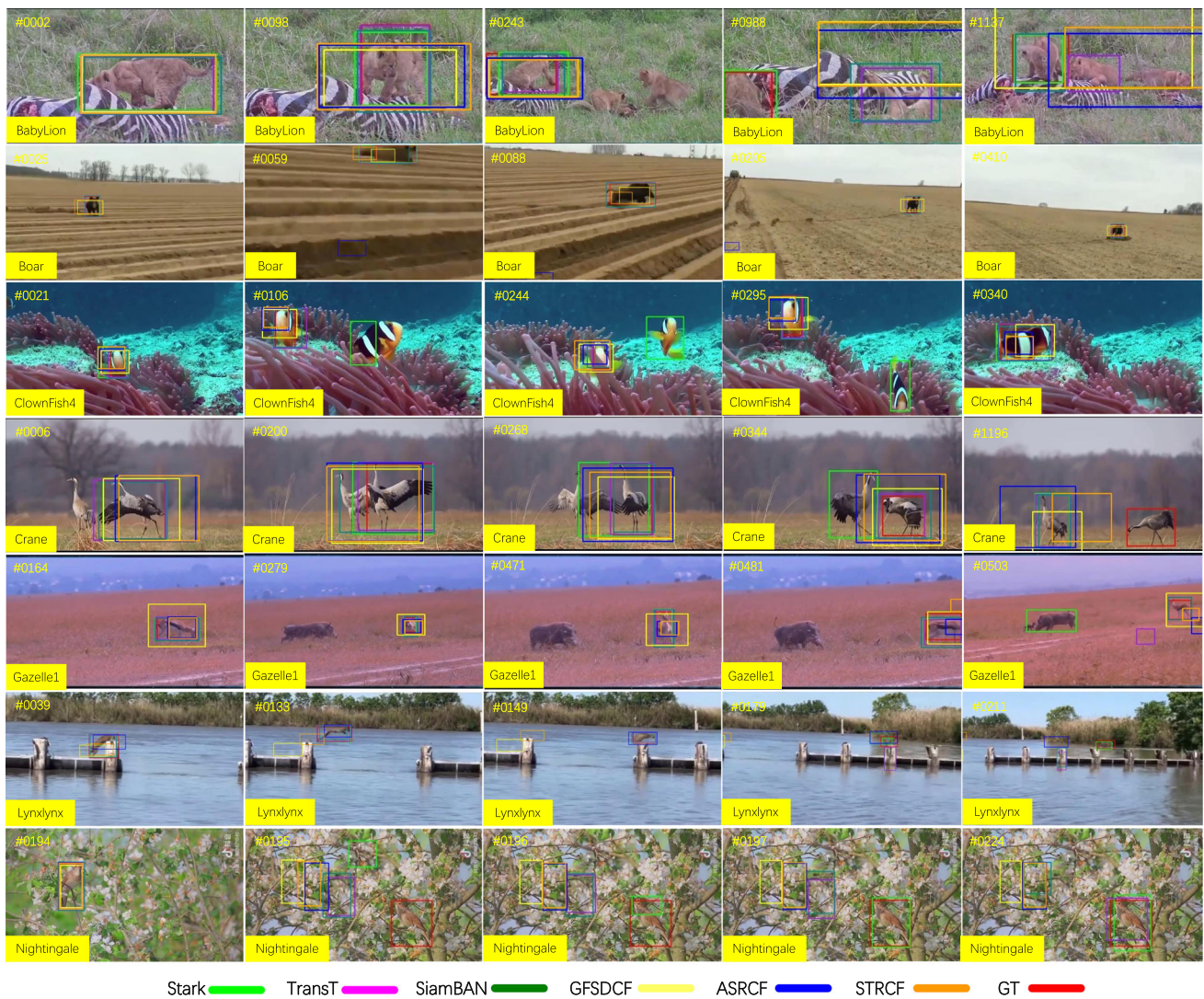


Fig. 9 Qualitative evaluation results

malized precision and success AUC (0.577, 0.512) are much lower than that on LaSOT dataset (0.770, 0.671 Yan et al. (2021)). For STMTrack, its AUC score on OTB100 is 0.719 Fu et al. (2021) which is much higher than that on WATB (0.486). It tells us that wild animal tracking is very challenging and is far from resolved. The declining performance of these tracking algorithms on WATB datasets occurs for various reasons, including missing training data, characteristic physical movements and constraints (e.g., biped, tetrapod, polyped), irregular or homogeneous skin texture, and limited resolution caused by video compression. Therefore, a series of innovations are required in the research of wild animal tracking, including wild animal appearance modeling, designing more robust and efficient deep network, building wild animal datasets for training, etc. Researchers from the computer vision community are encouraged to take part in the development of wild animal tracking.

6 Conclusion

In this paper, we introduce WATB, the first benchmark for wild animal tracking. WATB includes more than 200 video sequences with over 203,000 frames covering a wide range of animal species. To the best of our knowledge, WATB is the first benchmark released in the computer vision community for evaluating object tracking methods in wild animal tracking. Details of the construction of WATB is described so as to demonstrate our principles of building WATB. In order to understand the performance of existing tracking methods on WATB, we extensively evaluate 38 popular tracking methods, covering emerging Transformer based trackers, Siamese network based trackers, and traditional CF based trackers. We select CF based trackers with hand-crafted features, deep features and both to do extensive evaluation. The quantitative and qualitative results demonstrate that Transformer based

Table 7 English abbreviation control table

Abbreviation	Full
WATB	Wild animal tracking benchmark
CF	Correlation filter
OTB	Online tracking benchmark
NFS	Need for speed
LaSOT	Large-scale single object tracking
UAV123	Unmanned aerial vehicle 123
GOT10K	Generic object tracking 10K
TC128	Temple color 128
AMTSet	Abrupt motion tracking dataSet
TOTB	Transparent object tracking benchmark
WAMI	Wide area motion imagery
MOSSE	Minimum output sum of squared error filter
KCF	Kernelized correlation filter
STAPLE	Complementary learners for real-time tracking
SAMF	Scale adaptive kernel correlation filter tracker with feature integration
HOG	Histogram of oriented gradient
CN	Color naming
SRDCF	Spatially regularized discriminative correlation filter
BACF	Background aware correlation filter
STRCF	Spatial-temporal regularized correlation filter
ASRCF	Adaptive spatially-regularized correlation filter
GFSDCF	Group feature selection discriminative correlation filter
CNN	Convolutional neural network
MDNet	Multi-domain convolutional neural networks
SiamFC	Fully-convolutional Siamese network
SiamRPN	Siamese regional proposal network
SiamBAN	Siamese box adaptive network
SiamCAR	Siamese fully convolutional classification and regression
SiamAttn	Deformable siamese attention network

trackers show better performance in handling different challenging factors in wild animal tracking than Siamese network based trackers and CF based tracker. The results also tell us that Wild Animal Tracking is far from well resolved. We believe that WATB will ignite the interests in wild animal tracking and consequently promote the development of wild animal behavior analysis and wildlife preservation.

Supplementary Information

Some of the abbreviations and corresponding full names are shown in Table 7.

Acknowledgements This work was supported by the National Natural Science Foundation of China under Grants 61972068 and 61976042, Innovative Talents Program for LiaoningUniversities under Grant

LR2019020 and the Liaoning Revitalization Talents Program under Grant XLYC2007023.

Author Contributions FW and FS conceived this study. FW wrote the initial manuscript, and FS reviewed and edited it. The other four authors took part in the construction of WTAB. PC and FL are responsible for tracker evaluation. XW is responsible for building the project website.

Availability of Data and Materials The datasets generated during and/or analysed during the current study are available in the project website: <https://w-1995.github.io/>.

Declarations

Conflict of interest All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

References

- Bertinetto, L., Valmadre, J., Golodetz, S., Miksik, O., & Torr, P.H. (2016). Staple: Complementary learners for real-time tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1401–1409. IEEE, Las Vegas, USA. <https://doi.org/10.1109/CVPR.2016.156>
- Bertinetto, L., Valmadre, J., Henriques, J.F., Vedaldi, A., & Torr, P.H. (2016). Fully-convolutional siamese networks for object tracking. In: *European Conference on Computer Vision*, pp. 850–865. Springer, Amsterdam, Netherlands. https://doi.org/10.1007/978-3-319-48881-3_56
- Bhat, G., Danelljan, M., Gool, L.V., & Timofte, R. (2019). Learning discriminative model prediction for tracking. In: *IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea, pp. 6182–6191. <https://doi.org/10.1109/ICCV.2019.00628>
- Bolme, D.S., Beveridge, J.R., Draper, B.A., & Lui, Y.M. (2010). Visual object tracking using adaptive correlation filters. In: *International Conference on Computer Vision and Pattern Recognition*, pp. 2544–2550. IEEE, San Francisco, USA. <https://doi.org/10.1109/CVPR.2010.5539960>
- Cao, Z., Fu, C., Ye, J., Li, B., & Li, Y. (2021). Hift: Hierarchical feature transformer for aerial tracking. In: *IEEE/CVF Conference on Computer Vision*, pp. 15457–15466. IEEE/CVF, Montreal, QC, Canada. <https://doi.org/10.1109/ICCV48922.2021.01517>
- Chen, X., Yan, B., Zhu, J., Wang, D., Yang, X., & Lu, H. (2021). Transformer tracking. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 8126–8135. <https://doi.org/10.1109/CVPR46437.2021.00803>
- Chen, Z., Zhong, B., Li, G., Zhang, S., & Ji, R. (2020). Siamese box adaptive network for visual tracking. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6668–6677. IEEE/CVF, Seattle, WA, USA. <https://doi.org/10.1109/CVPR42600.2020.00670>
- Dai, K., Wang, D., Lu, H., Sun, C., & Li, J. (2019). Visual tracking via adaptive spatially-regularized correlation filters. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4670–4679. IEEE/CVF, Salt Lake City, USA. <https://doi.org/10.1109/CVPR.2019.00480>
- Danelljan, M., Bhat, G., Khan, F.S., & Felsberg, M. (2019). Atom: Accurate tracking by overlap maximization. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, pp. 4660–4669. <https://doi.org/10.1109/CVPR.2019.00479>
- Danelljan, M., Bhat, G., Shahbaz Khan, F., & Felsberg, M. (2017). Eco: Efficient convolution operators for tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6638–6646. IEEE, Honolulu, Hawaii. <https://doi.org/10.1109/CVPR.2017.733>
- Danelljan, M., Häger, G., Khan, F., & Felsberg, M. (2014). Accurate scale estimation for robust visual tracking. In: *British Machine Vision Conference*. Bmva Press, Nottingham, England. <https://doi.org/10.5244/C.28.65>
- Danelljan, M., Hager, G., Shahbaz Khan, F., & Felsberg, M. (2015). Learning spatially regularized correlation filters for visual tracking. In: *IEEE International Conference on Computer Vision*, pp. 4310–4318. IEEE, Santiago, Chile. <https://doi.org/10.1109/ICCV.2015.490>
- Danelljan, M., Häger, G., Khan, F. S., & Felsberg, M. (2016). Discriminative scale space tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(8), 1561–1575. <https://doi.org/10.1109/TPAMI.2016.2609928>
- Fan, H., Lin, L., Yang, F., Chu, P., Deng, G., Yu, S., Bai, H., Xu, Y., Liao, C., & Ling, H. (2019). Lasot: A high-quality benchmark for large-scale single object tracking. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5374–5383. IEEE/CVF, Long Beach, USA. <https://doi.org/10.1109/CVPR.2019.00552>
- Fan, H., Miththanathaya, H.A., Harshit, Rajan, S.R., Liu, X., Zhou, Z., Lin, Y., & Ling, H. (2021). Transparent object tracking benchmark. In: *International Conference on Computer Vision*, pp. 10734–10743. IEEE/CVF, Nashville, TN, USA. <https://doi.org/10.1109/ICCV48922.2021.01056>
- Feng, W., Han, R., Guo, Q., Zhu, J., & Wang, S. (2019). Dynamic saliency-aware regularization for correlation filter-based object tracking. *IEEE Transactions on Image Processing*, 28(7), 3232–3245. <https://doi.org/10.1109/TIP.2019.2895411>
- Fu, Z., Liu, Q., Fu, Z., & Wang, Y. (2021). Stmtrack: Template-free visual tracking with space-time memory networks. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 13774–13783. <https://doi.org/10.1109/CVPR46437.2021.01356>
- Fukunaga, T., Kubota, S., Oda, S., & Iwasaki, W. (2015). Grouptracker: Video tracking system for multiple animals under severe occlusion. *Computational Biology and Chemistry*, 57, 39–45. <https://doi.org/10.1016/j.compbiolchem.2015.02.006>
- Fu, C., Xu, J., Lin, F., Guo, F., Liu, T., & Zhang, Z. (2020). Object saliency-aware dual regularized correlation filter for real-time aerial tracking. *IEEE Transactions on Geoscience and Remote Sensing*, 58(12), 8940–8951. <https://doi.org/10.1109/TGRS.2020.2992301>
- Galoogahi, H.K., Fagg, A., Huang, C., Ramanan, D., & Lucey, S. (2017). Need for speed: A benchmark for higher frame rate object tracking. In: *International Conference on Computer Vision*, pp. 1134–1143. IEEE, Venice, Italy. <https://doi.org/10.1109/ICCV.2017.128>
- Guo, D., Shao, Y., Cui, Y., Wang, Z., Zhang, L., & Shen, C. (2021). Graph attention tracking. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, pp. 9543–9552. <https://doi.org/10.1109/CVPR46437.2021.00942>
- Guo, D., Wang, J., Cui, Y., Wang, Z., & Chen, S. (2020). Siamcar: Siamese fully convolutional classification and regression for visual tracking. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6269–6277. IEEE/CVF, Seattle, WA, USA. <https://doi.org/10.1109/CVPR42600.2020.00630>
- Haalck, L., Mangan, M., Webb, B., & Risse, B. (2020). Towards image-based animal tracking in natural environments using a freely moving camera. *Journal of Neuroscience Methods*, 330, 108455. <https://doi.org/10.1016/j.jneumeth.2019.108455>
- He, Z., Fan, Y., Zhuang, J., Dong, Y., & Bai, H. (2017). Correlation filters with weighted convolution responses. In: *IEEE International Conference on Computer Vision Workshop*, pp. 1992–2000. IEEE, Venice, Italy. <https://doi.org/10.1109/ICCVW.2017.233>
- Henriques, J.F., Caseiro, R., Martins, P., & Batista, J. (2012). Exploiting the circulant structure of tracking-by-detection with kernels. In: *European Conference on Computer Vision*, pp. 702–715. Springer, Florence, Italy. https://doi.org/10.1007/978-3-642-33765-9_50
- Henriques, J. F., Caseiro, R., Martins, P., & Batista, J. (2014). High-speed tracking with kernelized correlation filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(3), 583–596. <https://doi.org/10.1109/TPAMI.2014.2345390>
- Huang, Z., Fu, C., Li, Y., Lin, F., & Lu, P. (2019). Learning aberrance repressed correlation filters for real-time uav tracking. In: *IEEE/CVF International Conference on Computer Vision*, pp. 2891–2900. IEEE/CVF, Seoul, Korea(south). <https://doi.org/10.1109/ICCV.2019.00298>
- Huang, L., Zhao, X., & Huang, K. (2021). Got-10k: A large high-diversity benchmark for generic object tracking in the wild. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(5), 1562–1577. <https://doi.org/10.1109/TPAMI.2019.2957464>
- Kiani Galoogahi, H., Fagg, A., & Lucey, S. (2017). Learning background-aware correlation filters for visual tracking. In: *IEEE*

- International Conference on Computer Vision*, pp. 1135–1143. IEEE, Venice, Italy. <https://doi.org/10.1109/ICCV.2017.129>
- Li, Y., & Zhu, J. (2014). A scale adaptive kernel correlation filter tracker with feature integration. In: *European Conference on Computer Vision Workshop*, pp. 254–265. Springer, Zurich, Switzerland. https://doi.org/10.1007/978-3-319-16181-5_18
- Li, Y., Fu, C., Ding, F., Huang, Z., & Lu, G. (2020). Autotrack: Towards high-performance visual tracking for uav with automatic spatio-temporal regularization. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11923–11932. IEEE/CVF, Seattle, WA, USA. <https://doi.org/10.1109/CVPR42600.2020.01194>
- Li, F., Tian, C., Zuo, W., Zhang, L., & Yang, M.-H. (2018). Learning spatial-temporal regularized correlation filters for visual tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4904–4913. IEEE, Salt Lake City, USA. <https://doi.org/10.1109/CVPR.2018.00515>
- Li, B., Wu, W., Wang, Q., Zhang, F., Xing, J., & Yan, J. (2019). Siamrpn++: Evolution of siamese visual tracking with very deep networks. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4282–4291. IEEE/CVF, Long Beach, USA. <https://doi.org/10.1109/CVPR.2019.00441>
- Li, B., Yan, J., Wu, W., Zhu, Z., & Hu, X. (2018). High performance visual tracking with siamese region proposal network. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8971–8980. IEEE, Salt Lake City, USA. <https://doi.org/10.1109/CVPR.2018.00935>
- Liang, P., Blasch, E., & Ling, H. (2015). Encoding color information for visual tracking: Algorithms and benchmark. *IEEE Transactions on Image Processing*, 24(12), 5630–5644. <https://doi.org/10.1109/TIP.2015.2482905>.
- Li, C., Liang, X., Lu, Y., Zhao, N., & Tang, J. (2019). Rgb-t object tracking: Benchmark and baseline. *Pattern Recognition*, 96, 106977. <https://doi.org/10.1016/j.patcog.2019.106977>.
- Lin, Y., Cheng, S., Shen, J., & Pantic, M. (2019). Mobiface: a novel dataset for mobile face tracking in the wild. In: *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 1–8. <https://doi.org/10.1109/FG.2019.8756581>
- Liu, S., Liu, D., Srivastava, G., Polap, D., & Woźniak, M. (2021). Overview of correlation filter based algorithms in object tracking. *Complex and Intelligent Systems*, 7, 1895–1917. <https://doi.org/10.1007/s40747-020-00161-4>.
- Li, P., Wang, D., Wang, L., & Lu, H. (2018). Deep visual tracking: Review and experimental comparison. *Pattern Recognition*, 76, 323–338. <https://doi.org/10.1016/j.patcog.2017.11.007>.
- Li, F., Wu, X., Zuo, W., Zhang, D., & Zheng, L. (2020). Remove cosine window from correlation filter-based visual trackers: When and how. *IEEE Transactions on Image Processing*, 29, 7045–7060. <https://doi.org/10.1109/TIP.2020.2997521>.
- Lopez-Marcano, S., Jinks, E. L., Buelow, C. A., Brown, C. J., Wang, D., Kusy, B., et al. (2021). Automatic detection of fish and tracking of movement for ecology. *Ecology and Evolution*, 11, 8254–8263. <https://doi.org/10.1002/ece3.7656>.
- Lu, H., & Wang, D. (2019). *Online Visual Tracking*. Singapore: Springer.
- Ma, C., Huang, J.-B., Yang, X., & Yang, M.-H. (2015). Hierarchical convolutional features for visual tracking. In: *IEEE International Conference on Computer Vision*, pp. 3074–3082. IEEE, Santiago, Chile <https://doi.org/10.1109/ICCV.2015.352>
- Mathis, A., Mamidanna, P., Cury, K.M., Abe, T., N., M.V., W., M.M., & Bethge, M. (2018). Deeplabcut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience* 21, 1281–1289. <https://doi.org/10.1038/s41593-018-0209-y>
- Mueller, M., Smith, N., & Ghanem, B. (2016). A benchmark and simulator for uav tracking. In: *European Conference on Computer Vision*, pp. 445–461. Springer, Amsterdam, The Netherlands. https://doi.org/10.1007/978-3-319-46448-0_27
- Nam, H., & Han, B. (2016). Learning multi-domain convolutional neural networks for visual tracking. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4293–4302. IEEE, Las Vegas, USA. <https://doi.org/10.1109/CVPR.2016.465>
- Norouzzadeh, M. S., Nguyen, A., Kosmala, M., Swanson, A., & Palmer, M. S. (2018). Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. *Proceedings of the National Academy of Sciences of the United States of America*, 115(25), 5716–5725. <https://doi.org/10.1073/pnas.1719367115>.
- Ravoor, P. C., & Sudarshan, T. S. B. (2020). Deep learning methods for multi-species animal re-identification and tracking—a survey. *Computer Science Review*, 38, 100289. <https://doi.org/10.1016/j.cosrev.2020.100289>.
- Risse, B., Mangan, M., Del Pero, L., & Webb, B. (2017). Visual tracking of small animals in cluttered natural environments using a freely moving camera. In: *2017 IEEE International Conference on Computer Vision Workshops*, pp. 2840–2849. IEEE/CVF, Venice, Italy. <https://doi.org/10.1109/ICCVW.2017.335>
- Risse, B., Mangan, M., Del Pero, K., & Webb, B. (2017). Visual tracking of small animals in cluttered natural environments using a freely moving camera. In: *International Conference on Computer Vision Workshop*, pp. 2840–2849. <https://doi.org/10.1016/j.anbehav.2016.12.005>
- Shen, J., Zafeiriou, S., Chrysos, G.G., Kossaiji, J., Tziropoulos, G., & Pantic, M. (2015). The first facial landmark tracking-in-the-wild challenge: benchmark and results. In: *International Conference on Computer Vision Workshop*, pp. 1003–1011. IEEE, Santiago, Chile. <https://doi.org/10.1109/ICCVW.2015.132>
- Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., et al. (2022). Perspectives in machine learning for wildlife conservation. *Nature Communication*, 13(792), 1–15. <https://doi.org/10.1038/s41467-022-27980-y>.
- Valletta, J. J., Torney, C., Kings, M., Thornton, A., & Madden, J. (2017). Applications of machine learning in animal behaviour studies. *Animal Behaviour*, 124, 203–220. <https://doi.org/10.1016/j.anbehav.2016.12.005>.
- Valmadre, J., Bertinetto, L., Henriques, J.F., Tao, R., Vedaldi, A., Smeulders, A.W.M., Torr, P.H.S., & Gavves, E. (2018). Long-term tracking in the wild: A benchmark. In: *European Conference on Computer Vision*, pp. 692–707. Springer, Munich, Germany. https://doi.org/10.1007/978-3-030-01219-9_41
- Van der Zande, L. E., Guzhva, O., & Rodenburg, T. B. (2021). Individual detection and tracking of group housed pigs in their home pen using computer vision. *Frontiers in Animal Science*, 2, 669312. <https://doi.org/10.3389/fanim.2021.669312>.
- Wang, L., Ouyang, W., Wang, X., & Lu, H. (2015). Visual tracking with fully convolutional networks. In: *IEEE International Conference on Computer Vision*, pp. 3119–3127. IEEE, Santiago, Chile. <https://doi.org/10.1109/ICCV.2015.357>
- Wang, Q., Zhang, L., Bertinetto, L., Hu, W., & Torr, P.H.S. (2019). Fast online object tracking and segmentation: A unifying approach. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, pp. 1328–1338. <https://doi.org/10.1109/CVPR.2019.00142>
- Wang, F., Wang, C., Yin, S., He, J., Sun, F., & Zhang, J. (2022). Amtset: A benchmark for abrupt motion tracking. *Multimedia Tools and Applications*, 81, 4711–4734. <https://doi.org/10.1007/s11042-021-10947-4>.
- Weinstein, B. G. (2018). A computer vision for animal ecology. *Journal of Animal Ecology*, 87, 533–545. <https://doi.org/10.1111/1365-2656.12780>.
- Wu, Y., Lim, J., & Yang, M.-H. (2013). Online object tracking: A benchmark. In: *IEEE Conference on Computer Vision and Pat-*

- tern Recognition*, pp. 2411–2418. IEEE, Portland, USA. <https://doi.org/10.1109/CVPR.2013.312>
- Wu, Y., Lim, J., & Yang, M. H. (2015). Object tracking benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(9), 1834–1848. <https://doi.org/10.1109/TPAMI.2014.2388226>.
- Xiong, F., Zhou, J., & Qian, Y. (2020). Material based object tracking in hyperspectral videos. *IEEE Transactions on Image Processing*, 29, 3719–3733. <https://doi.org/10.1109/TIP.2020.2965302>.
- Xu, T., Feng, Z., Wu, X., & Kittler, J. (2019). Joint group feature selection and discriminative filter learning for robust visual object tracking. In: *IEEE/CVF International Conference on Computer Vision*, pp. 7950–7960. IEEE/CVF, Seoul, Korea. <https://doi.org/10.1109/ICCV.2019.00804>
- Xu, T., Feng, Z.-H., Wu, X.-J., & Kittler, J. (2019). Learning adaptive discriminative correlation filters via temporal consistency preserving spatial feature selection for robust visual object tracking. *IEEE Transactions on Image Processing*, 28(11), 5596–5609. <https://doi.org/10.1109/TIP.2019.2919201>.
- Xu, T., Feng, Z., Wu, X., & Kittler, J. (2021). Adaptive channel selection for robust visual object tracking with discriminative correlation filters. *International Journal of Computer Vision*, 129, 1359–1375. <https://doi.org/10.1007/s11263-021-01435-1>.
- Yan, B., Peng, H., Fu, J., Wang, D., & Lu, H. (2021). Learning spatio-temporal transformer for visual tracking. In: *IEEE/CVF International Conference on Computer Vision*, Montreal, QC, Canada, pp. 10448–10457. <https://doi.org/10.1109/ICCV48922.2021.01028>
- Ye, J., Fu, C., Lin, F., Ding, F., An, S., & Lu, G. (2022). Multi-regularized correlation filter for uav tracking and self-localization. *IEEE Transactions on Industrial Electronics*, 69(6), 6004–6014. <https://doi.org/10.1109/TIE.2021.3088366>.
- Yu, Y.a. (2020). Deformable siamese attention networks for visual object tracking. In: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6728–6737. IEEE/CVF, Seattle, WA, USA. <https://doi.org/10.1109/CVPR42600.2020.00676>
- Zheng, G., Fu, C., Ye, J., Lin, F., & Ding, F. (2021). Mutation sensitive correlation filter for real-time uav tracking with adaptive hybrid label. In: *IEEE International Conference on Robotics and Automation*, pp. 503–509. IEEE, Xi'an, China. <https://doi.org/10.1109/ICRA48506.2021.9561931>
- Zuo, W., Wu, X., Lin, L., Zhang, L., & Yang, M.-H. (2018). Learning support correlation filters for visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(5), 1158–1172. <https://doi.org/10.1109/TPAMI.2018.2829180>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.