

A SMALL TARGET VISUAL TRACKING METHOD FOR UNMANNED AERIAL VEHICLE PLATFORM UNDER CONVOLUTIONAL NEURAL NETWORK

Jiya Tia¹, Yong Zhang¹, Chao Wang¹, Pingping Xiao¹, Shuyan Zhang¹, Baiyan Xu¹
¹Electric Information College of Changchun Guanghua University, Changchun, 130031, China.
Email: tianyiyalaos@yeah.net

Abstract - In order to explore the easy loss of tracking target in the realization of Unmanned Aerial Vehicle (UAV) platform visual target tracking, the convolutional neural network algorithm is proposed to improve the existing vision tracking method of small targets on UAV platform so as to be better applied to UAV flight platform. Firstly, in order to solve the real-time requirement of small target detection of UAV platform, the convolutional neural network target detection model is improved. A lightweight detection model combining Mobile Net V2 with SSD algorithm is built. On this basis, the focal loss function is added for modification and optimization. The accurate detection of small targets can be guaranteed under the condition that the model operation amount is small. Then, based on the TLD algorithm, visual significance target detection is introduced into the TLD framework, and a high-precision video target tracking algorithm with local re-detection capability is proposed. The tracking results are effectively corrected to reduce the error enlargement in long-term tracking. Finally, the new algorithm and the original algorithm are compared qualitatively and quantitatively through simulation experiments. The results show that compared with the original model, the improved model has a small amount of computation, and the accuracy of the algorithm is significantly improved. Not only that, the new algorithm has a stronger ability of heavy detection, which improves the robustness of the algorithm and successfully realizes the detection and tracking of small targets under the platform of UAV. Therefore, the lightweight convolutional neural network algorithm proposed in this research lays the application foundation for target detection of UAV platform and further promotes the development of target tracking technology in UAV video.

Keywords: Convolutional Neural Network; UAV; Target Tracking; Target Detection.

1. Introduction

Drones have long been used in the military. However, in order to meet the needs of military strike, intelligence gathering and long-distance flight, the size of traditional military UAV is often designed to be very large. Coupled with the high cost of manufacturing drones, they have not been widely used for a long time. With the continuous development of science and technology, the reduction of UAV cost and the emergence of intelligent applications, more and more UAV have been applied to a variety of environments and industries. In addition, small portable UAV has become the focus of current consumers, through which a large number of aerial video resources can be obtained.

Tracking and detecting the video images to obtain the objects of interest has become a widespread application demand in many industries. Therefore, the realization of UAV with various intelligent functions has become one of the important research fields of many scholars [1].

Video target detection and target tracking have been a research hotspot in the field of computer vision. But in the use of a small platform like a drone, the targets in the field of view are too small to be recognized, which adds many difficulties to target detection and tracking. In addition, due to the small size of UAV platform, the performance of airborne equipment is generally low due to the consideration of load, endurance, and other aspects. However, many existing algorithms are too complex to be directly transplanted and used. In the process of visual tracking of small targets on UAV platform, the moving image is often blurred due to the fast-moving speed of the target. Moreover, during the movement of the target, changes in ambient light and the degree of occlusion affect the tracking performance of the UAV to the target [2,3]. Nowadays, convolutional neural networks help more and more computer vision tasks, such as target detection and scene segmentation, achieving a leap in accuracy and robustness.

Although convolutional neural network has made great achievements in target detection, most

scholars agree that convolutional neural network models such as VGG have too much computation and serious parameter redundancy. For example, under the platform of UAV, these algorithms need to be further improved to meet people's needs due to the limitation of its load, the interference caused by the small proportion of tracking targets, and low resolution [4]. Therefore, the convolutional neural network has a very large research space in compression model parameters, control operation complexity, and so on. In recent years, some miniaturized convolutional neural network models have been proposed, including the Squeeze Net model with low memory footprint, the Mobile Net lightweight model designed for Mobile terminals, and the Shuffle Net cell architecture suitable for low-power devices. The birth of these models indicates that it has become a development trend in the field of deep learning to build a lightweight convolutional neural network that is convenient for mobile terminals [5]. Due to extensive research and mature performance of convolutional neural networks and deep learning technologies, these typical technologies have excellent presentation effects in many computer vision tasks. Therefore, in the field of target tracking, it can be tried to combine research with deep learning technology to gradually realize that the accuracy and robustness of target tracking algorithm are not affected by interference factors.

To sum up, several classical algorithms are adopted as the starting point of improvement to determine the direction of improvement. In order to realize autonomous target tracking, the UAV platform must first find the interested target in the field of vision, so the target detection algorithm is essential. Therefore, based on the target detection algorithm of convolutional neural network, the strategy combining Mobile Net V2 and SSD algorithm is first selected. In order to further improve the ability of the algorithm in this research to detect small objects, the focal loss function is introduced to optimize the design. Secondly, based on the original TLD algorithm model, a significance target detector module based on cellular automata is added. Combined with significance detection, the tracking accuracy and robustness of the algorithm are improved. By comparing the algorithm with the original TED algorithm, the superiority of the proposed algorithm is proved.

2. Literature Review

In recent years, with the continuous development of technology, new convolutional neural network models emerge one after another, with increasingly higher accuracy, but also with the doubling growth of network depth, number of feature maps, and computation [6]. In practical application, the platform with small capacity and low power consumption is difficult to graft the model.

Cheng et al. (2017) proposed a quantization method based on convolutional neural network in their study, and obtained the convolutional network after acceleration and compression, while the consumption of memory and storage is much lower [7]. Laga et al. (2017) used the performance characteristics of SSD to accelerate the classification process of the algorithm and proposed a new sorting algorithm MONTRES. MONTRES relies on SSD performance models. When the ratio of file size to main memory size is high, MONTRES is better than the latest algorithm, because it can shorten the sorting execution time of TPC-H dataset by more than 30% [8]. In the aspect of target tracking, Kalal first proposed that the TLD algorithm has strong robustness and timeliness, so it is widely known, improved and utilized. This algorithm has an online learning mechanism that can constantly update the detector and tracker, which not only improves the anti-occlusion capability of the whole algorithm, but also lengthens the time of target tracking [9]. Jin and Liu (2016) proposed a new accelerated TLD algorithm to improve the running speed of the algorithm. In the detection module, the gray prediction model is used to estimate the target position and set the detection area. The results show that the algorithm runs fast on the premise of ensuring accuracy [10]. Zhang et al. (2017) proposed a reliable point assignment algorithm based on wavelet transform to solve the problems of object occlusion and appearance deformation in visual tracking. The proposed RPA and Kalman filter are integrated into the tracking detection framework to improve the tracking accuracy [11]. Liu et al. (2019) coordinated the resource and bandwidth issues based on Zynq heterogeneous platform and roofline model, and designed the convolutional neural network accelerator to accelerate the standard convolution sum depth separable product at a high hardware resource rate [12].

To sum up, convolutional neural network has been widely applied to vision tasks. Lightweight convolutional neural model has gradually become a research hotspot at home and abroad, but relatively few studies have applied it to UAV flight platforms with limited performance. Therefore, a lightweight small target detection algorithm based on convolutional neural network is proposed, which can not only detect small targets at a long distance, but also reduce the algorithm complexity, which is more suitable for the use of UAV platform. Secondly, the significance target detection algorithm based on cellular automata is combined with the TLD algorithm model to construct a video target tracking algorithm with local re-detection capability.

The optimization algorithm proposed in this research lays a foundation for the follow-up research of both tracking accuracy and computation.

3. Methodology

3.1 A lightweight Mobile Net architecture based on convolutional neural network

Mobile Net model takes the depth separation convolution as the main algorithm. The independent little convolution kernel convolution is done on each channel. Then, 1×1 point convolution is used for channel fusion. Supposing the number of channels in the convolution is W_{in} , and the number of channels in the output is W_{out} , only symmetric convolution is considered, the size of the convolution kernel is h , and the dpi of the convolution feature graph is $c \times i$. Then the traditional equation for the number of convolution operations is as follows.

$$N = h^2 ciW_{in}W_{out} \tag{1}$$

The number of operations required by deep convolutional neural network is as follows.

$$N = h^2 ciW_{in} + ciW_{in}W_{out} \tag{2}$$

Thus, it can be concluded that compared with traditional convolution, deeply separated convolution can greatly reduce the computation, and its equation is expressed as follows.

$$\frac{h^2 ciW_{in} + ciW_{in}W_{out}}{h^2 ciW_{in}W_{out}} = \frac{1}{W_{out}} + \frac{1}{h^2} \tag{3}$$

When the convolution kernel of 3×3 is modified, the deep convolutional neural network only needs 0.1 times the operation amount of the traditional convolution.

3.2 A small target feature extraction model based on convolutional neural network

Among the target detection algorithms of the depth convolutional neural network with one-stage as the typical model, the SSD algorithm model is the most representative. The specific testing process of SSD algorithm is: firstly, the target image is transported into VGG-16 network for convolution feature extraction; secondly, the multi-scale feature graph calculated from the main part of VGG-16 network is transferred into the detection network, and the target is classified and located at the same time; thirdly, there will be differences in the scale of different layers of the multi-layer feature map. The box size is set in the form of a scale. The scaling calculation equation is as follows.

$$S_k = S_{min} + \frac{S_{max} - S_{min}}{m - 1} (k - 1) \tag{4}$$

In Eq. 4, S_{min} is the first layer, which is defaulted as 0.2; S_{max} is the last layer, which is defaulted as 0.9; the middle layer is analogized, and the default number of layers in SSD300 is 6.

For the original SSD target detection model, the backbone network of Mobile Net V2 classification

model is used to replace the original VGG-16 network.

The new model is nine tenths less computational than the original. To this extent, in order to further reduce the complexity of detection network operation, the running speed should be improved. In this model, the default block diagrams of 1:3 and 3:1 with different proportions in the original SSD model are removed. Other architectures are basically consistent with the original SSD, and the superimposed model architecture of Mobile Net V2 and SSD detection model is shown in Figure 1.

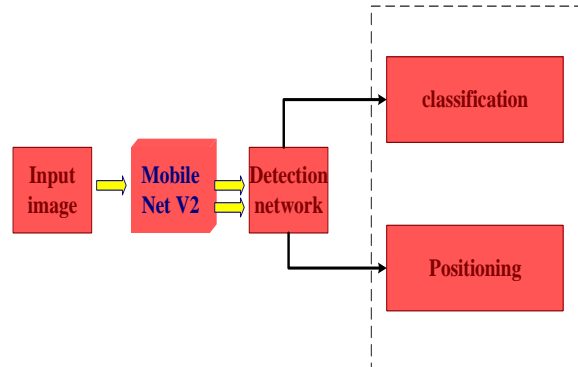


Figure 1: Target detection model of Mobile Net V2 based on convolutional neural network

3.3 Model adjustment based on the focal loss function

In order to adjust the proportion principle of positive and negative samples, which is similar to the hard-negative sample mining strategy in SSD, the weight on the loss function is adjusted, and the equation is expressed as follows.

$$L = -\alpha \log(p_i, \hat{p}_i) \tag{5}$$

On this basis, in order to enhance the sensitivity of the whole algorithm to complex samples, a confidence related variable is introduced to adjust its influence, and its equation is expressed as follows.

$$L = -\alpha(1 - \hat{p}_i)^\gamma \log(p_i, \hat{p}_i) \tag{6}$$

After being tested for different values, α is taken as 0.75 in the research; γ is defaulted as 2.

Small targets detected by UAVs have few high-level features, which are difficult to distinguish from large targets. After the lightweight algorithm model ensures the running speed, the feature extraction ability is relatively weak. Therefore, an improved focal loss function is proposed in this research. By increasing the super-parameter and the weight of complex samples on the focal loss function, the sensitivity of the network to small targets is improved. The equation of the model focal loss function proposed in this research is as follows.

$$L_{c,r}(\{p_i\}, \{t_i\}) = L_c + \lambda L_r \tag{7}$$

L_c is the classification loss, L_r is the positioning loss; $L_{c,r}$ represents a combined loss.

The equation after introducing the loss function into the classification loss is as follows.

$$L_c = \frac{1}{N_c} \sum_{i=1}^{N_c} \left[-\alpha (1 - \hat{p}_i)^\gamma \log(p_i, \hat{p}_i) \right] \quad (8)$$

The positioning loss equations are as follows.

$$L_r = \sum smooth_{L1}(t_i - \hat{t}_i) \quad (9)$$

$$smooth(x) = \begin{cases} 0.5x^2 & \text{if } |x| < 1 \\ |x| - 0.5 & \text{otherwise} \end{cases} \quad (10)$$

In this research, the model is built and trained under the TensorFlow flow. The specific flow diagram is shown in Figure 2.

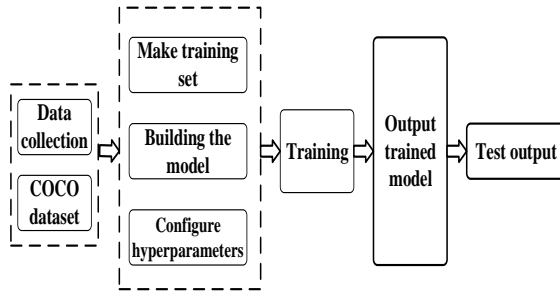


Figure 2: The model training flow chart.

Eq. 11 shows the validation expression of small target detection algorithm effect of UAV platform.

$$P = \frac{TP}{TP + FP}$$

$$R = \frac{TP}{TP + FN} \quad (11)$$

$$F = \frac{2 \times P \times R}{P + R}$$

In Eq. 11, P is expressed as precision, R is expressed as recall rate, TP and FP are expressed as real cases and false positive cases respectively, and FN is expressed as false negative cases. F is the value of comprehensive evaluation.

3.4 Significance detection of TLD video target tracking

Target tracking is like a person's eyes, which is to achieve a strong ability to capture. It is necessary to estimate the tracking target in advance, judge the trajectory of the target, and infer the motion state of the target; at the same time, rapid redetection is essential. Once the target is out of sight, adjustments can be made. Under the platform of UAV, the small target is always in the state of rapid movement, causing multiple problems such as the change of perspective and target occlusion. If these problems are not solved, they may easily lead to the loss of tracking targets and failure of tracking results. One of the ways to solve this problem is to improve the ability of the target tracking algorithm. Therefore,

visual significance target detection is introduced into the TLD framework in this research, and local redetection capability is applied to high-precision video target tracking algorithm.

The TLD algorithm is mainly divided into four modules: tracker, detector, learner, and synthesizer. Due to the limitation of detection accuracy and the inflexible matching mechanism, the tracking target is often lost in the process of fast moving. Therefore, based on the original TLD framework, the improved algorithm in this research adds a significance target detector module based on cellular automata, as shown in Figure 3. Through the local redetection proposed in this research combined with the TLD algorithm, it is expected to be able to accurately track the position of the box and improve the target tracking ability.

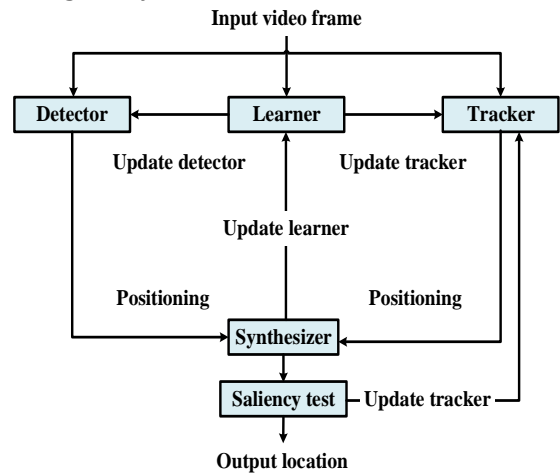


Figure 3: Improved algorithm flow.

Based on the output characteristics of the original algorithm, the algorithm in this research locates in the center of the bounding box and the equation of its center point is expressed as follows.

$$center_x = x + \frac{width}{2} \quad (12)$$

$$center_y = y + \frac{height}{2} \quad (13)$$

The search scope of the surrounding box is increased, and its equation is expressed as follows.

$$x_{new} = center_x - \frac{\lambda}{2} width$$

$$y_{new} = center_y - \frac{\lambda}{2} height \quad (14)$$

$$width_{new} = \lambda \times width$$

$$height_{new} = \lambda \times height$$

In Eq. 14, λ is the scale factor, and the value range is $2 \leq \lambda \leq 3$. The larger λ is, the larger the recheck box, the stronger the recheck ability. But as λ gets bigger, the calculations get more complicated; if λ is too large, it causes multiple

detection targets to appear in the search area, affecting the accuracy of significance detection. The image in the expanded bounding box is sent to the visual significance target detector based on cellular automata. The value range of the significance graph is 0~255. The threshold of 128 is used as the segmentation point, and significance greater than this value is circled by a new range box, which is the significance box. Taking the coordinates of the significance box as the result output, the central point of the significance box is calculated, and compared with the central point of the original tracking box. The offset equation is as follows.

$$\begin{aligned} \Delta x &= \text{SaliencyCenter}_x - \text{Center}_x \\ \Delta y &= \text{SaliencyCenter}_y - \text{Center}_y \end{aligned} \quad (15)$$

The scaled offset is transmitted to the target tracker in the TLD for real-time update of the target location.

The scaling process is as follows:

$$\begin{aligned} \text{NewCenter}_x &= \text{Center}_x + \frac{\Delta x}{\lambda} \\ \text{NewCenter}_y &= \text{Center}_y + \frac{\Delta y}{\lambda} \end{aligned} \quad (16)$$

An error in the significance result causes the trace target to change, and to avoid this, the offset is scaled. However, in the process of scaling, it is easy to see that the search box is too large, leading to detection of multiple tracking targets and affecting the accuracy of the results. The scale factor is used to scale inversely. If the interference of multiple detection targets occurs occasionally, the results of significance detection will be affected, and all tracking boxes will also be mis-located. In order to improve the robustness of the tracking algorithm, the confidence degree of the significance box is increased on the basis of the real-time update of the target position. In other words, the confidence degree of an adaptive significance box is replaced by the original scale factor of inverse scaling.

Its equation is expressed as follows:

$$\begin{aligned} \text{NewCenter}_x &= \text{Center}_x + \alpha \times \hat{\eta} \times \Delta x \\ \text{NewCenter}_y &= \text{Center}_y + \alpha \times \hat{\eta} \times \Delta y \end{aligned} \quad (17)$$

$\hat{\eta}$ is the confidence of the significance box; α is the super-parameter, and the value in this research is 1.2. The idealized significance graph detection value is 255, and the background significance value is 0, that is, the trace target is detected exactly.

At this point, the significance graph shows the characteristic of maximizing the total variance.

The normalized variance of the significance graph is used as the confidence of the importance box.

The specific calculation method is as follows.

$$\eta = \sqrt{\sum_{i=0}^{\text{width}_{new}} \sum_{j=0}^{\text{height}_{new}} (s_{ij} - \bar{s})^2} \quad (18)$$

$$\hat{\eta} = \frac{\eta}{128 \sqrt{\text{width}_{new} \times \text{height}_{new}}} \quad (19)$$

In Eq. 18, s_{ij} is the significance value of the corresponding pixel point; \bar{s} is the mean of significance values of all pixel points.

The tracking accuracy can be determined by the distance between the target center point and the tracking center point. The longer the distance, the greater the tracking error, the shorter the distance, the more accurate the tracking result. The specific calculation process is as follows.

$$GT_x = \frac{x_1 + x_2 + x_3 + x_4}{4} \quad (20)$$

$$GT_y = \frac{y_1 + y_2 + y_3 + y_4}{4} \quad (21)$$

$$d = \sqrt{(\bar{x} - GT_x)^2 + (\bar{y} - GT_y)^2} \quad (22)$$

The ground truth information marked in VOT data set is expressed as x_1, x_2, x_3, x_4 and y_1, y_2, y_3, y_4 . \bar{x} and \bar{y} represent the position of the center point of the algorithm output tracking box; d represents the distance between the target center point and the tracking center point.

4. Results

4.1 Small target detection algorithms comparison results

In order to verify the feasibility of the small object extraction detection algorithm designed in this study, the InceptionV2+SSD, MobileNetV1+SSD, and MobileNetV2+SSD algorithms are selected for comparison with the algorithm in this research.

Figure 4 shows its average accuracy and the comprehensive evaluation F value.

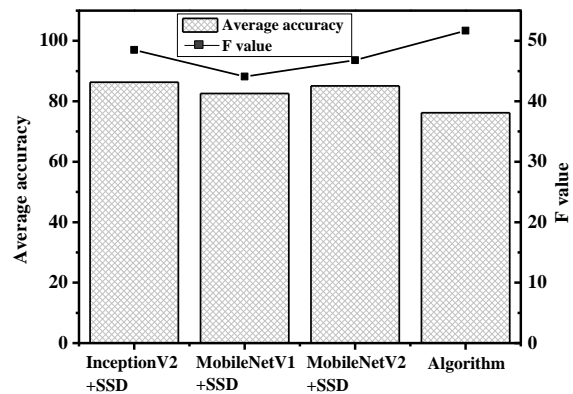


Figure 4: The test results comparison of each algorithm.

By comparing the average recall rate of target detection algorithm, InceptionV2+SSD algorithm reaches 33.6%, MobileNetV1+SSD algorithm reaches 30.8%, MobileNetV2+SSD algorithm reaches 32.4%, and the algorithm in this research reaches 39.8%.

Therefore, it can be concluded from the figure and data that the average accuracy of the algorithm in this research is about 8% lower than that of the InceptionV2+SSD algorithm with the highest accuracy, but the recall rate and the comprehensive evaluation F value are both higher than the other three algorithms. Moreover, the parameter usage of the improved algorithm is only about half of that of the InceptionV2+SSD model.

In addition, the detection rate of the target is also greatly improved.

To sum up, the optimization algorithm proposed in this research has obvious advantages over other target detection algorithms in terms of recall rate, and the target detection rate and comprehensive evaluation F value are also higher in the same type of lightweight target detection algorithm.

4.2 Simulation experiment of small target visual tracking on UAV platform

An improved TLD algorithm based on convolutional neural network is proposed to solve the problem of accurate visual tracking of small targets in UAV platform. The proposed algorithm is compared with the original TLD in UAV video sequence. In order to verify the effectiveness of the algorithm proposed in this research in UAV platform, the experiment is based on the open source computer vision library for simulation comparison. The video sequence in VOT2017 video data set is selected as the test data set to test the proposed algorithm. The hardware environment tested is a computer with a Tntel i5-3470 GPU and 8GB RAM, and the software environment is a Windows operating system.

Figure 5 shows the experimental tracking comparison results in the simulation experiment, where yellow is the original TLD algorithm and red is the algorithm in this research.

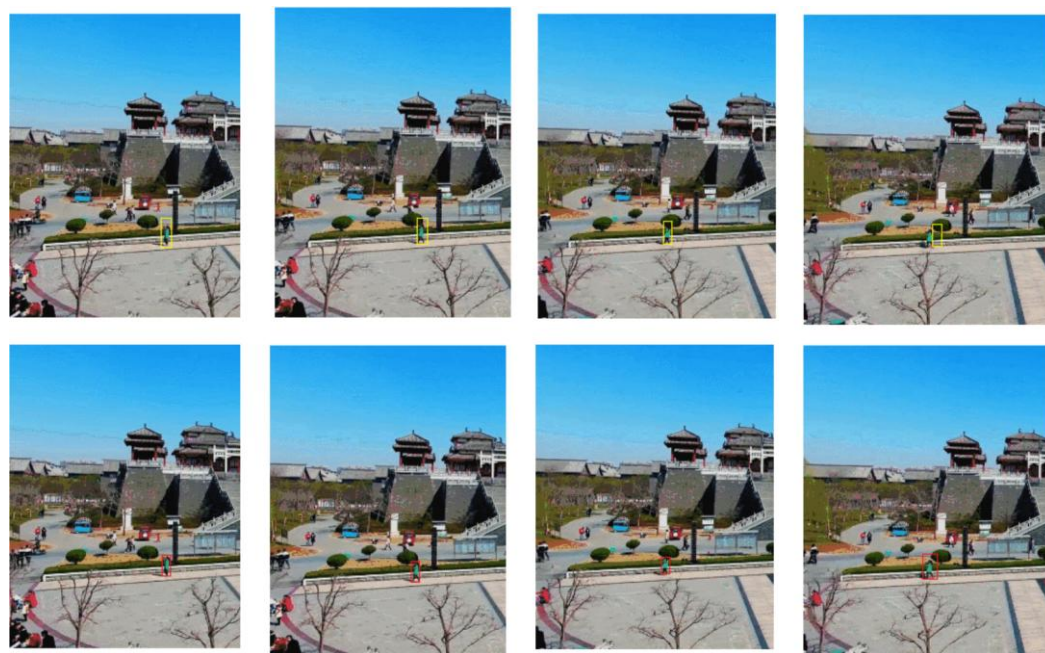


Figure 5: The tracking results comparison of the original TLD algorithm and the algorithm in this research.

As shown in the figure, the tracking box of the original algorithm can also keep up with the target at the beginning, but the tracking box is larger than the target, indicating that the tracking accuracy is slightly less. However, the tracking box proposed in this research is closely surrounded by the target at the beginning, and the target tracking precision is relatively higher. In the later stage of the experiment, it is found that the tracking box is larger than the original target in both algorithms.

However, by comprehensive comparison, the algorithm results proposed in this research are closer to the tracking target.

4.3 Quantitative comparison

Taking Pedestrian2 VOT2015 data set sequence for an example, Figure 6 shows the error curve.

The red curve is the error result of the algorithm proposed in this research, and the black curve is the error result of the original TLD algorithm.

As shown in Figure 6, the center position error of the algorithm in this research is always kept at a low level compared with the original algorithm, indicating that it has a better tracking effect in the target tracking process.

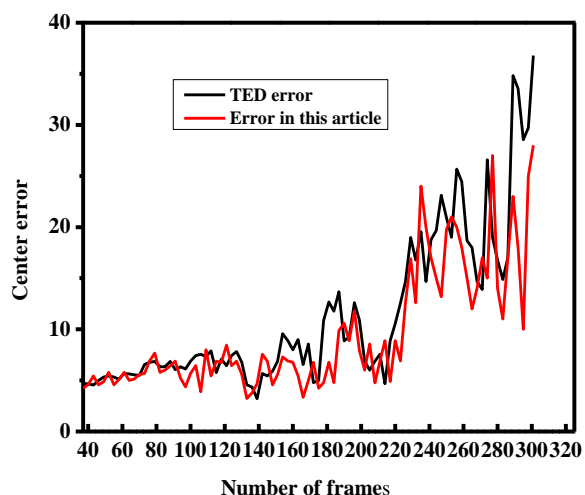


Figure 6: Curve of error variation.

In this research, some difficult sequences are selected for a large number of comparison experiments.

Table 1 shows the average error comparison results to verify the re-detection capability of the new algorithm.

Table 1 Data of tracking results of each sequence

Video sequence	The original TLD algorithm	The algorithm in the research
Jogging1	6.13	3.98
Drone-across	6.45	3.55
Motocross1	38.89	15.96
Pedestrian2	6.93	3.54

Table 1 shows that for long-term tracking target pedestrian 2 sequence.

The average error of the proposed a lgorithm is reduced by about half compared with the original algorithm; for the drone across sequence with significant deformation, the algorithm proposed in this research can not only reduce the average error, but also track more frames, with relatively strong redetection ability; for motocross1 sequences with strong deformation, the algorithm proposed in this research tracks the same frame value as the original TLD algorithm, but the average error of the algorithm proposed in this research is smaller; similarly, for Jogging1 sequence, the average error of the proposed algorithm is smaller.

Therefore, compared with the original algorithm, the algorithm proposed in this research has stronger redetection capability and higher accuracy, so the robustness of the algorithm is also improved.

5. Discussion

As an important part of computer vision, target tracking technology provides important information value for motion force estimation, video analysis, and other tasks. Many typical algorithms come into being, such as Struck, Mean Shift, and so on. Although these algorithms can be operated in an ideal environment, if the target extraction features are not obvious and the diversity of the detection environment changes, the performance of these algorithms will be poor. With the emergence of convolutional neural network and the great success achieved in the field of computer vision, the improved algorithm based on convolutional neural network has become a hot research topic in recent years, greatly improving the algorithm performance of research tasks [13].

For example, the combination of target tracking algorithm and convolutional neural network makes the algorithm far more accurate than the traditional algorithm. However, due to the complexity of convolutional neural network, its tracking speed in visual tracking task is slowed down, which affects the tracking effect. Liu et al. (2019) proposed a UAV target tracking algorithm based on adaptive depth network, and designed a third-order adaptive depth network to extract target features. In this network, the H, S, and I channels of images are studied by principal component analysis, and the obtained feature vectors are input into the network for hierarchical convolution, which optimizes the network structure, improves the convergence speed and accuracy of the network, and effectively avoids the decline in tracking accuracy caused by the interference of complex factors [14]. Jin et al. (2019) proposed a multi-target tracking algorithm in a clutter environment based on the problems existing in traditional radar target tracking.

Assisted by doppler information and IPDA-UKF algorithm framework, a multi-dimensional correlated wave-gate based on target position and velocity information is established. Through a large number of measurements and effective screening of multidimensional information, the correlation between distance and velocity is utilized to successfully decompose a multi-target data into multiple single-target data correlation problems. It reduces the complexity of multi-target data association in clutter environment and improves the accuracy of target tracking [15]. In this study, the balance between computation and tracking accuracy is taken into account.

A lightweight small target detection model is proposed to adapt to the UAV platform, and the focal loss function is introduced to adjust and correct it.

While ensuring the accuracy of target detection, the complexity of operation is reduced.

The target tracking method is based on the original TED model, and the significance detection algorithm of cellular automata is added to improve the visual target tracking performance of UAV.

6. Conclusions

In order to explore a small target visual tracking method suitable for UAV platform, according to the difficulty of applying video small target tracking technology to UAV platform, the algorithm is improved from two aspects: target detection and target tracking. A lightweight detection model based on convolutional neural network and the improved TLD structure model are proposed and compared on the measured UAV aerial photography data set.

The results show that the improved target detection algorithm proposed in this research has obvious advantages in target detection rate compared with the original algorithm, and the comprehensive evaluation F value is also improved in the same type of lightweight target detection algorithm. Since the TLD video target tracking algorithm combined with significance detection adopts local redetection, the increase in computation is not much, but the accuracy and robustness are obviously improved.

An in-depth study is made on the visual detection and tracking of small targets on UAV platform, and its improved algorithm enhances the tracking performance of UAV object recognition, activity analysis, and other aspects, providing reference experience for future research on small target detection, long-term tracking, and other related technologies.

However, in the lightweight small target detection model, although the model built by the introduced focal loss function has a better detection effect for small targets, the problem of missing detection for large targets occurs.

Therefore, the present loss function has not achieved the desired effect, and further research is needed in this regard.

References

- [1] Tsai Y. H. (2019) Vision-based collision avoidance for unmanned aerial vehicles by recurrent neural networks. *International Journal of Computer and Information Engineering*, 13(4), 196-200.
- [2] Yadav N., Binay U. (2017) Comparative study of object detection algorithms. *International Research Journal of Engineering and Technology (IRJET)*, 4(11), 586-591.
- [3] Guo K., Sui L., Qiu J, et al. (2017) Angel-Eye: A complete design flow for mapping CNN onto embedded FPGA. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 37(1), 35-47.
- [4] Wang X., Dong X., Kong X, et al. (2017) Drogue detection for autonomous aerial refueling based on convolutional neural networks. *Chinese Journal of Aeronautics*, 30(1), 380-390.
- [5] Dorj B., Kang D. C., Lee D. J. (2016) Vision based target tracking control of a multirotor drone using raspberry pi embedded module. *Advanced Science Letters*, 22(10), 3043-3046.
- [6] Pugliese L. D. P., Guerriero F., Zorbas D, et al. (2016) Modelling the mobile target covering problem using flying drones. *Optimization Letters*, 10(5), 1021-1052.
- [7] Cheng J., Wu J., Leng C, et al. (2017) Quantized CNN: A unified approach to accelerate and compress convolutional networks. *IEEE transactions on neural networks and learning systems*, 29(10), 4730-4743.
- [8] Laga A., Boukhobza J., Singhoff F, et al. (2017) Montres: merge on-the-run external sorting algorithm for large data volumes on SSD based storage systems. *IEEE Transactions on Computers*, 66(10), 1689-1702.
- [9] Shao H., Shen J., Zhang Z, et al. (2018) Research and analysis of video image target tracking algorithm based on significance. *International Journal of High-Performance Systems Architecture*, 8(1-2), 82-93.
- [10] Jin Z., Liu C. (2016) Accelerated TLD algorithm and its application in multiple target tracking. *Computer Syst Appl*, 25(6), 196-201.
- [11] Zhang R., Deng T., Wang G, et al. (2017) A robust object tracking framework based on a reliable point assignment algorithm. *Frontiers of Information Technology & Electronic Engineering*, 18(4), 545-558.
- [12] Liu B., Zou D., Feng L, et al. (2019) An FPGA-based CNN accelerator integrating depth wise separable convolution. *Electronics*, 8(3), 281-282.
- [13] Cheng Y., Wang D., Zhou P, et al. (2018) Model compression and acceleration for deep neural networks: The principles, progress, and challenges. *IEEE Signal Processing Magazine*, 35(1), 126-136.
- [14] Liu F., Wang H. J., Huang G. W, et al. (2019) UAV target tracking algorithm based on adaptive depth network. *Acta Aeronautica ET Astronautica Sinica*, 40(3), 174-183.
- [15] Jin B., Li C., Guo J, et al. (2019) Multi-target tracking in clutter aided by doppler information. *Electronic Science*. 48(4), 511-517.