Theory and Practice of Science and Technology
2025, VOL. 6, NO. 9, 141-144
DOI: 10.47297/taposatWSP2633-456928.20250609

Survey of UAV Visual tracking Technology based on Convolutional Neural Networks

Yage Xie

Zhengzhou University of Light Industry, Zhengzhou, Henan, 450000, China

ABSTRACT
In this paper, the unmanned aerial vehicle (UAV) based on convolution neural network visual tracking technology has carried on the comprehensive review, including its development course, research status and development trend in the future. Through combing and analyzing therelevant literature, the advantages and disadvantages of different algorithms in UAV visual tracking are summarized, and some research directions worthy of in-depth discussion are proposed. This paper also summarizes and analyzes the current challenges and problems, and discusses how to further improve the performance and stability of UAV visual tracking technology. Finally, the application prospect of UAV visual tracking technology in military and civilian fields is prospected, which provides reference and inspiration for related research and development work.

KEYWORDS
Unmanned aerial vehicle; Visual tracking; Convolutional neural network; Autonomous flight; Target recognition; Flight safety; Deep learning technology; Image processing

1 Introduction

1.1 Research background

At present, with the gradual popularization and development of UAV technology, UAV plays a pivotal role in various fields. Drones visual tracking technology as one of the key technology of autonomous navigation and target tracking, received the attention of the researchers. We interpret tracking as the process of locating an object of interest defined in the current frame in the subsequent frames of a video. Namely, "location", "object of interest", and "subsequent frame", correspond to the three sub-problems in the basic principle of tracking: candidate box generation, feature representation/extraction, and decision making. Through this complete logical chain, visual object tracking is realized [1].

Through the UAV visual tracking technology based on convolutional neural network, the UAV can realize real-time tracking and recognition of the target, and then realize more intelligent and accurate flight operation. This technology can not only be applied to target tracking and surveillance in the military field, but also can be widely used in aerial photography, security patrol, agricultural monitoring and other fields.

At the same time, with the rapid increase in the number of Uavs and the diversification of UAV application scenarios, the demand for UAV visual tracking technology is also increasing. Researchers need to continuously improve and optimize this technology to adapt to the actual needs of different fields, improve the autonomous flight and target tracking capabilities of Uavs, and promote the further development and application of UAV technology.

1.2 Research significance

UAV visual tracking technology is of great significance for improving the autonomy of UAV. Through machine vision perception and tracking technology, unmanned aerial vehicle (UAV) can realize real-time monitoring and tracking of targets, so as to realize autonomous flight and task execution. This can not only reduce the work burden of the operator, but also improve the accuracy and safety of the flight. UAV visual tracking technology can also be applied to various fields, such as military reconnaissance, disaster monitoring, and environmental protection. In the military field, UAV visual tracking technology can be used for target recognition and attack to improve operational efficiency and accuracy. In disaster monitoring, UAVs can monitor disaster areas in real time and provide important information to support rescue work. In environmental protection, UAVs can be used to monitor the number of wild animals and the situation of protected areas to help protect the ecological environment.

In general, the research on UAV visual tracking technology is of great significance for improving the autonomous judgment ability of UAV, improving flight safety and improving mission execution efficiency, and will bring new opportunities and challenges to the development of UAV field.

* **Corresponding Author:** Yage Xie, 2759839131@qq.com

2 Overview of UAV technology

2.1 Development history of UAVs

From the introduction of the concept of UAV in the early 20th century, to the military application during World War II, and then to the extensive application in the commercial and scientific research fields in recent years, the development of UAV technology has progressed by leaps and bounds. With the wide application of advanced sensors, cutting-edge communication technology and flight control system, UAVs show more and more bright application prospects in the military and civilian fields.

In the military field, UAVs are widely used in reconnaissance, reconnaissance, target tracking, strike and other tasks, which improve the efficiency of combat and the accuracy of information acquisition. In the civilian field, UAVs are used in agriculture, environmental monitoring, logistics and distribution, power inspection and other fields, which greatly improve work efficiency and safety.

With the continuous development of deep learning technology, the visual tracking technology of UAV based on convolutional neural network has also been greatly developed [2]. This technology can realize the automatic identification, tracking and positioning of the target, which provides important technical support for the autonomous flight of UAV. In the future, UAV technology is expected to show greater application potential in more fields, bringing more convenience and help to human society.

2.2 Overview of UAV vision technology

A UAV is an unmanned aerial vehicle that performs various tasks through pre-set procedures or remote operation. Drones vision technology refers to using the camera sensor devices, such as real-time acquisition, processing and analysis of the surrounding environment of image information, in order to realize the target object detection, tracking and recognition. The technology in military surveillance, disaster relief, agricultural plant protection and other fields has a broad application value. The UAV visual tracking technology based on convolutional neural network combines deep learning technology with UAV visual technology, and realizes the automatic tracking and recognition of the target object by training a neural network model. This technology can greatly improve the unmanned aerial vehicle (UAV) in the target tracking, video monitoring, accuracy and efficiency of the task, for the application of unmanned aerial vehicle (UAV) in various fields provides a powerful technical support.

In the future, with the continuous development and improvement of artificial intelligence technology, the visual tracking technology of UAV based on convolutional neural network will be more mature and popular, which will bring more possibilities and opportunities for the application of UAV technology.

3 The basis of convolutional neural network

3.1 Principle of convolutional neural network

Convolution Neural Network (Convolutional Neural Network, CNN) is a kind of convolution computation of deep learning Neural Network, is mainly used for image recognition and classification task. Its basic principle is through the Convolution (Input), pooling layer (Convolution) and the connection layer (Fully Connected) building complex network structure, in order to extract the image feature and classification and processing. In the CNN, convolution layer through the filter to the corresponding convolution of the input image, to extract the characteristics of different levels. Pooling layer is used to reduce the image characteristics of spatial dimensions, thus reducing the number of arguments and enhance the robustness of the model. The fully connected layer is responsible for mapping the features to the output labels, and finally completes the classification task.

CNN's advantage is can image characteristics of autonomic learning, no manual design feature extractor, and through multi-level feature extraction, using the convolution and pooling layer stack, better improve the generalization ability and accuracy of the model and the expression ability for complex data. CNN also can be Shared by convolution kernels and weights of sharing mechanism to reduce the model and quantity, improve the computational efficiency.

In the uav visual tracking technology, CNN is widely used in target detection, target tracking and recognition tasks such as scene. Its excellent feature extraction ability and classification accuracy make it possible for unmanned aerial vehicle (UAV) to monitor and track targets more accurately, which has broad application prospects.

3.2 Application of convolutional neural network in image processing

Convolutional Neural Network (CNN) is a kind of feed forward neural network, which is composed of a convolution layer, a pooling layer and a fully connected layer. In the field of image processing, CNN has become a very effective technique, which is widely used in image recognition and classification, object detection, scene understanding and other tasks.

The use of CNN in image processing is mainly reflected in its ability to extract and recognize image features. By convolution layer and pooling of pooling layers, CNN can self-study features in the image, so as to realize the identification and classification of the complex image data. Compared with traditional image processing methods, CNN

has better generalization ability and robustness, and performs better in dealing with large-scale image data.

In the UAV visual tracking technology, image processing method based on CNN is becoming more and more important. Using CNN to extract image features and combining with tracking algorithms to achieve target tracking has achieved remarkable results in the field of UAV. In the future, with the continuous progress of deep learning technology, the visual tracking technology of UAV based on CNN will be further improved, providing more possibilities and application scenarios for autonomous flight and target tracking of UAV.

3.3 Target tracking algorithm based on convolutional neural network

3.3.1 Single object tracking algorithm based on convolutional neural network

The single object tracking algorithm based on Convolutional Neural Network (CNN) is a method to track selected objects in videos. The core key of the algorithm is to use CNN to detect and track the target. Target tracking based on deep features uses deep convolutional features to replace traditional hand-crafted features, which aims to improve the feature expression ability. After extracting features, correlation filtering is generally used for target discrimination, and target tracking is performed frame-by-frame [3].

In each frame in the video, the use of CNN model (such as Faster - R - CNN, YOLO or SSD) to detect the existence of the target and determine its position and size. The trained model, to identify specific targets from the image, in the first frame box in the boundary of the target template and return box (bounding box) coordinates [4]. Once the target is detected, the algorithm can extract target established by CNN convolution layer characteristics, including color, texture, shape, etc. As a result, the target has a consistent representation between different frames, which is helpful for the subsequent tracking process. In the tracking stage, the target template is obtained in the first frame, and in the subsequent frames, the system will find the regions with similar features extracted in the previous frame by using the algorithm, so as to realize the continuous tracking of the target. This usually utilizes some form of matching algorithm, such as the Hungarian algorithm, to find the best matching target location. Target tracking is achieved by repeating this process over and over. Next, the object in the current frame is associated with the object in the previous frame. Here, an algorithm such as Kalman filter, particle filter, or mean shift can be used to estimate the trajectory of the object. It helps to predict the position of the object in the next frame and maintain continuous tracking of the object between multiple frames.

In order to improve the accuracy of tracking, the algorithm will also update the CNN model by using online learning or semi-supervised learning methods. The target images acquired during the tracking process or other images from similar scenes are used as new training data. The algorithm is adapted to the appearance changes of the target and other interference factors in the scene. The single object tracking algorithm based on convolutional neural network can achieve fast and accurate tracking of moving objects in videos, and can achieve good results in videos captured by either a still camera or a moving camera. This algorithm in video surveillance, human-computer interaction, autopilot, and other fields has a broad application prospect.

3.3.2 Multi-object tracking algorithm based on convolutional neural network

Multi-object tracking (MOT) is a more challenging machine vision task. The purpose of online multi-object tracking is to determine the specific position of the object through the image information of each previous frame and the current frame, and always maintain the identity of the object, and obtain the process of each object's respective trajectory [5].

The multi-object tracking algorithm based on convolutional neural network can be divided into several parts: object detection, feature extraction, feature matching, trajectory updating, creating new trajectories, and object tracking. Common Convolutional Neural Network (CNN) architectures, such as Faster R-CNN, YOLO, and SSD, are used to extract the location and category information of the object in the selected image. For target is detected, convolution neural network further to extract the features, including the target motion, appearance and other characteristics. These features will be used as the basis for subsequent target tracking. In consecutive video frames, the algorithm to extract the target characteristic and the existing tracking trajectory for matching. If an object is successfully matched to a trajectory in the current frame, the trajectory information will be updated, including position, speed and direction. However, if an object is not matched to any of the existing trajectories in the current frame, a new trajectory will be created. After the above steps, the system outputs the tracking trajectories of each object, which can be used for subsequent tasks such as object behavior analysis, scene understanding, etc.

The multi-object tracking algorithm based on convolutional neural network is an important research hotspot in the field of computer vision today, and has very important practical application value. Its advantage is that it can effectively deal with various changes of the target, including scale change, occlusion, illumination change, etc. However, this algorithm requires high computational resources and faces no small challenges.

4 Overview of UAV visual tracking technology

4.1 Development history of UAV visual tracking technology

UAV visual tracking technology has experienced the development process from traditional methods to convolutional

neural network [6]. Traditional methods mainly include techniques based on feature matching and motion estimation. Although target tracking can be achieved to a certain extent, limited by factors such as illumination and background changes, the stability and robustness need to be improved.

With the development of deep learning, unmanned aerial vehicles (UAVs) based on convolution neural network visual tracking technology has gradually become the research hot spot [7]. Through the end-to-end learning of neural network, image features can be effectively extracted to achieve more accurate target recognition and tracking. At the same time, deep learning can also deal with complex environmental changes and improve the robustness and stability of the system.

At present, unmanned aerial vehicles (UAVs) based on convolution neural network visual tracking technology has achieved many breakthroughs, such as target detection, target tracking, long time tracking, etc. However, since deep learning requires a large amount of data and computing resources, how to improve the real-time performance and reduce the cost is still one of the focuses of future research. In the future study, can further explore multimodal fusion, migration study, such as technology, improve the performance and reliability of the UAV visual tracking technology.

4.2 Research status of UAV visual tracking technology based on convolutional neural network

At present, UAV visual tracking technology based on convolutional neural network has achieved a series of important research results [8]. Among them, some studies realize the accurate tracking of UAV targets by designing different network structures and algorithms, such as the accurate identification and tracking of targets in complex environments, and realize a more efficient UAV visual tracking system. Some studies have also explored how to effectively combine deep learning technology and traditional visual tracking methods to improve the accuracy and robustness of UAV in target tracking. These research results show a wide range of application prospects in multiple application scenarios, such as the application of UAVs in surveillance, search and rescue, military reconnaissance and other fields. By applying convolutional neural network technology to UAV visual tracking, it can make UAV more intelligent and efficient to perform tasks, and further expand the application range and capabilities of UAV.

5 Conclusion and Prospect

UAV visual tracking technology is one of the research hotspots in the field of UAV. Through the deep learning technology based on convolutional neural network, it realizes the real-time tracking and recognition of the target, so that the UAV can perform the task more intelligently and accurately. This technology has great application potential in military, aerial photography, security, agriculture and other fields, which can improve the autonomy, flight safety and task execution efficiency of UAV, and promote the further development and application of UAV technology.

The visual tracking technology of UAV based on convolutional neural network automatically learns image features and extracts target information to achieve accurate tracking and positioning of targets in complex environments with high robustness and real-time performance. With the continuous development and improvement of deep learning technology, the UAV visual tracking technology based on convolutional neural network will be more mature and popular, which brings more possibilities and opportunities for the application of UAV technology. At present, the research in this field mainly focuses on improving the accuracy and speed of the model, and solving the problems of real-time performance and cost. In the future, multimodal fusion, transfer learning and other technologies can be further explored to improve the performance and reliability of UAV visual tracking technology. Through continuous optimization and improvement, UAV visual tracking technology will bring new opportunities and challenges to the development of the UAV field, and bring more convenience and help to human society.

References

[1] Fiaz M ,Mahmood A ,Javed S , et al.Handcrafted and Deep Trackers: Recent Visual Object Tracking Approaches and Trends.[J].ACM Comput. Surv.,2019,52(2):43:1-43:44.

[2] Duan Haoqi. Research on Driving Action Recognition Based on Convolutional Neural Network [D]. Tutor: LyU Zhao; Zhang Chao. Anhui University, 2022.

[3] ZHAO Qinyan. Research on video single object tracking technology based on convolutional neural Network [D]. China University of Petroleum (East China), 2023.

[4] Huang Zan. Single Object Tracking Algorithm Based on Convolutional Neural Network [D]. Beijing Institute of Technology,2021.

[5] Wang Xianhao. Research on pedestrian detection and multi-target tracking algorithm based on convolutional neural Network [D]. Chongqing University of Posts and Telecommunications,2021.

[6] Yang Xiaoyun. Research on visual location recognition method based on convolutional neural network [D]. Tutor: Dai Tianhong. Northeast Forestry University, 2021.

[7] Wang Yi. Research on aerial image object detection algorithm based on convolutional neural network [D]. Supervisor: Yang Youlong. Xidian University, 2021.

[8] Zhang Ruiting. Research on helicopter flight state recognition technology based on one-dimensional convolutional Neural Network [D]. Tutor: Chen Xinyun; Xio.