Detection and Tracking of Small Targets in Remote Sensing Images

Net B - Group 4

KONG ZIYANG*
Beijing Institute of Technology

CHEN JIAYI
Chongqing University

MIAO TIANSHI
Southwest Jiaotong University

CHENG ZHAN
Beijing University of Posts and Telecommunications

DU GUANCHEN
Shantou University

GONG RUONAN
Southwest Jiaotong University

* Correspondence: 2284557925@qq.com

I. Abstract

In the context of remote sensing imagery, the accurate detection and tracking of small targets pose significant challenges. This study delves into the intricate task of identifying and monitoring these minute targets within such images. To address this, we embark on a comprehensive investigation of existing methodologies, with the intention to not only understand their intricacies but also to replicate their results. In addition, our project takes a step further by introducing a visually intuitive interface. This interface serves as a valuable tool, enhancing the usability of our work by providing a platform for efficient small target detection and tracking. Through this combined effort of research, replication, and innovation, our study aims to contribute to the advancement of target detection and tracking in remote sensing images.

Keywords: object detection, object tracking, remote sensing

II. Project Overview

Remote sensing imagery refers to Earth surface images acquired through remote platforms such as satellites and aircraft, used for studying and analyzing surface features and their changes [1]. Among the significant applications of remote sensing technology, target detection within these images stands as a cornerstone, holding irreplaceable importance in scientific research, resource management, environmental monitoring, and related domains. Nonetheless, the extensive coverage of Earth's surface in remote sensing imagery leads to a complex interplay of various scales and resolutions of objects, necessitating particular attention to smaller targets such as vehicles and architectural details [2]. These smaller targets, due to their relatively reduced scales, often pose challenges in accurate detection and identification when situated within intricate backgrounds.

In response to the issues arising from the diminutive size of targets, intricate backgrounds, and the difficulty in feature extraction within remote sensing images, this study embarks upon a comprehensive exploration of detection and tracking methodologies specifically tailored to the identification of small targets within remote sensing images. The components of the project include the following.

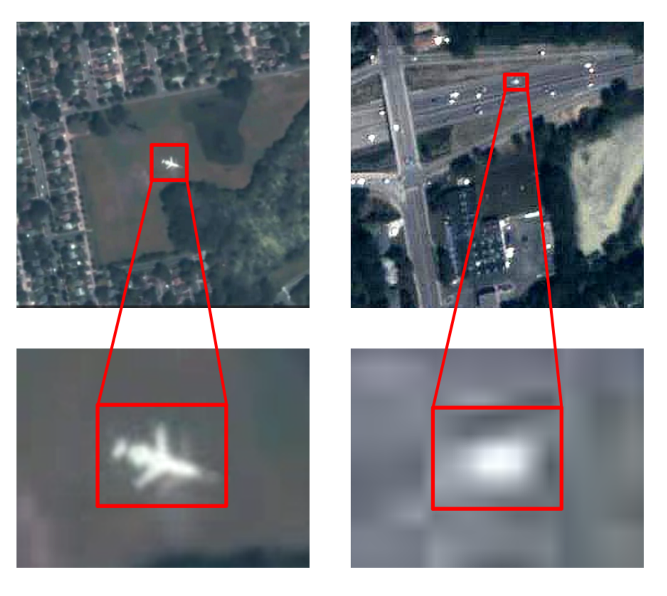


Fig. 1. Small Targets in Remote Sensing Imagery.

A. In-depth Survey

We undertook an in-depth survey of deep learning-driven small target detection and tracking methodologies and harnessing cutting-edge advancements in target detection and tracking technologies for small-scale objects embedded within remote sensing images

B. YOLOv5 and Bytetrack

We used improved YOLOv5 and Bytetrack algorithm to achieve detection and tracking of small targets in remote sensing images

C. Visual interface

We designed a visual interface to Display our object detection and tracking results.

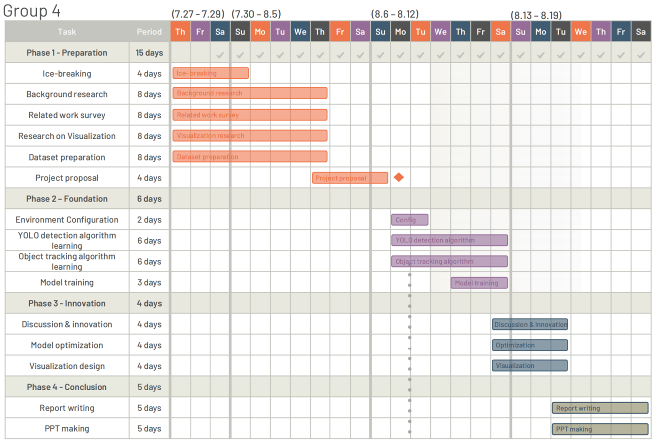
III. PROJECT BACKGROUND

Remote sensing imagery, as a crucial means of acquiring information about the Earth's surface, possesses distinctive attributes such as extensive coverage, high resolution, and multispectral data. Remote sensing image object detection, serving as a pivotal remote sensing technique, plays a significant role in applications like environmental monitoring, resource management, and disaster assessment. For instance, by identifying vehicles and structures within urban areas, it becomes possible to achieve traffic management and urban planning. Monitoring vegetation conditions in agricultural fields enables crop surveillance and prediction [3].

Nonetheless, due to the presence of numerous small-scale objects on the Earth's surface, such as vehicles, buildings, and vegetation, these small targets often exhibit challenges in remote sensing images, including weak features, low contrast, and noise interference [4]. These issues make their accurate detection and recognition challenging, and traditional object detection algorithms struggle to achieve the desired results. Given the difficulties in detecting small objects in remote sensing images, the significance of studying their detection and tracking becomes increasingly apparent. Effective small object detection and tracking techniques can enhance the information extraction capability of remote sensing images, offering more precise data support for environmental monitoring and resource management. Furthermore, with the advancement of artificial intelligence and deep learning, researchers have the opportunity to explore more advanced approaches to overcome the challenges of small object detection in remote sensing images, such as deep learning-based object detection models, thereby further enhancing detection accuracy and robustness.

IV. DIVISION OF ROLES

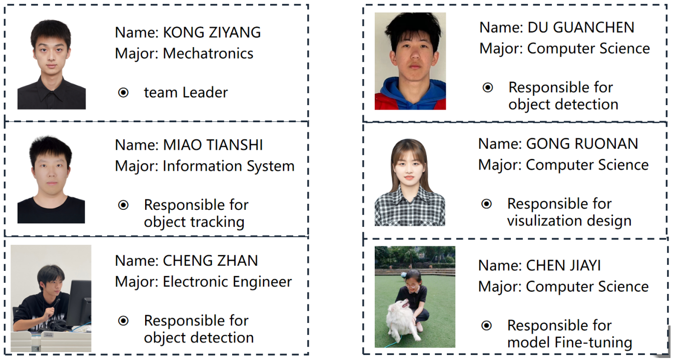
The project is divided into four phases, which are the preparation phase, the foundation phase, the Innovation phase, and the conclusion phase. The timeline for each phase is shown in the following Gantt chart.



Task	Period
Phase 1 - Preparation	15 days
Ice-breaking	4 days
Background research	8 days
Related work survey	8 days
Research on Visualization	8 days
Dataset preparation	8 days
Project proposal	4 days
Phase 2 – Foundation	6 days
Environment Configuration	2 days
YOLO detection algorithm learning	6 days
Object tracking algorithm learning	6 days
Model training	3 days
Phase 3 - Innovation	4 days
Discussion & innovation	4 days
Model optimization	4 days
Visualization design	4 days
Phase 4 - Conclusion	5 days
Report writing	5 days
PPT making	5 days

Fig. 2. Project Timeline.

Based on the specialties of the project members, the division of labor scheme of the project is shown in Fig.3.



Name: KONG ZIYANG
Major: Mechatronics
⦿ team Leader

Name: MIAO TIANSHI
Major: Information System
⦿ Responsible for object tracking

Name: CHENG ZHAN
Major: Electronic Engineer
⦿ Responsible for object detection

Name: DU GUANCHEN
Major: Computer Science
⦿ Responsible for object detection

Name: GONG RUONAN
Major: Computer Science
⦿ Responsible for visulization design

Name: CHEN JIAYI
Major: Computer Science
⦿ Responsible for model Fine-tuning

Fig. 3. Task Allocation.

V. CHALLENGES AND SOLUTIONS

A. Object Detection

1) Challenges:

We have many difficulties with small target detection. In the investigations of Liu Xiaobo [5] we can see the following. 1) Super large image size: Remote sensing images have super large image size and coverage area. Typical detection algorithms target small image sizes, which are difficult to directly apply in the field of remote sensing. At the same time, the background in the remote sensing image accounts for a large proportion, the target area is small, the typical detection method treats each area equally, and the calculation efficiency is extremely low.

2) Large change of direction: the remote sensing image is taken from the aerial perspective, the scene is a top view, and the target is distributed in the scene at multiple angles, most algorithms are not highly adaptable to angles, and they are not robust enough when dealing with multi-directional problems. In addition, the classic horizontal box positioning method is not compact enough and the positioning is not fine enough when positioning multi-directional targets.

3) Large scale of small targets: The proportion of small targets in remote sensing images is large. Small targets are easily lost due to feature degradation in the existing detection algorithms, resulting in missed detection. The method of small target detection has not been well defined in conventional natural image detection, and the field of remote sensing has increased the difficulty of detection.

4) Dense target distribution: There are large-scale and densely distributed targets in remote sensing images, and mutual interference between targets easily occurs, resulting in large positioning errors, and the problems of missed detection and false detection also occur very easily.

2) Related Work:

A good attempt to solve the above problems has been proposed in terms of an Improved High Precision Aircraft Target Detection Method of YOLT [6]. To achieve high-precision aircraft target detection in the background of remote sensing images, this paper proposes an improved YOLT aircraft target detection method to solve the problem of large target size variation and small targets. Firstly, focusing on the shortage of aircraft target datasets of remote sensing images, we constructed an aircraft detection dataset RSAD (Remote Sensing Aircraft Detection) and innovatively adopted a variety of data enhancement methods. Then, we analyzed the image of aircraft targets deeply under the same resolution, size distribution, and the range of small target pixel information in a network design, used the targeted context connection module, designed the layer number and the location of the pooling layer, which cause aircraft of different sizes to have better adaptability, and completed the receptive field in the form of dilated convolution of ascension. Finally, in the post-processing stage, the traditional non-maximum suppression method was improved to improve the detection accuracy again.

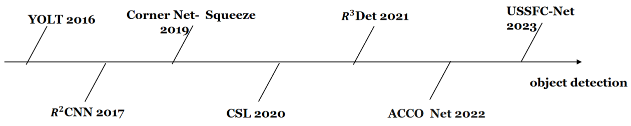


Fig. 4. Related Work of Object Detection.

3) Solution:

Remote sensing target detection algorithms can be roughly divided into three categories: detection methods based on classical pattern recognition, detection methods based on traditional machine learning, and detection methods based on deep learning. YOLT (You Only Look Twice), is a deep learning algorithm for object detection. Its core idea is to improve the accuracy of object detection through two-stage processing. YOLT can solve a lot of problems. Firstly, it uses a two-stage process to narrow the range of the target position, thereby improving the accuracy of target positioning. The first phase identifies areas that may contain targets through global observation, while the second phase performs more granular classification and localization in the local validation phase. Secondly, YOLT is also designed to handle targets at different scales. By introducing multiple receptive fields in convolutional neural networks, YOLTs can detect targets at different scales, from small to large. Because of the two-stage process, YOLT is able to filter out some false detections and improve the reliability of object detection. In the first stage, YOLT performs a preliminary classification of the image through sliding window technology, thus avoiding complex processing of the entire image. This method can improve the processing efficiency of large-scale images. In our opinion, the YOLT principle is simple. In the first stage, YOLT performs a preliminary classification of the image through sliding window technology, thus avoiding complex processing of the entire image. This method can improve the processing efficiency of large-scale images. In the second phase, YOLT extracts features from areas that may contain targets in the first stage for more granular classification and localization. This helps reduce false detections and improves the accuracy of target positioning. With this two-stage processing, YOLT is better able to adapt to targets at different scales and improve overall inspection performance.

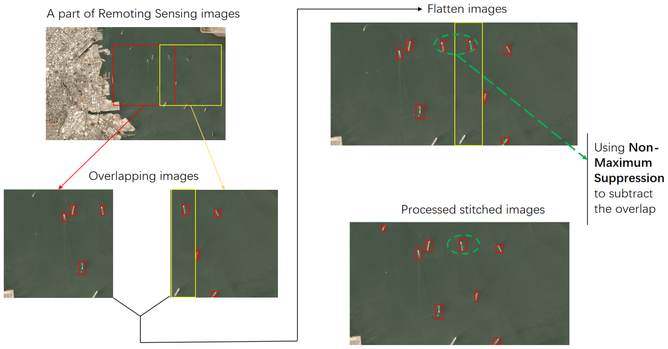

A part of Remoting Sensing images
Flatten images
Overlapping images
Using Non-Maximum Suppression to subtract the overlap
Processed stitched images

Fig. 5. Improvement of YOLO.

The steps in using YOLT can be summarized in this way. 1) User-defined bin sizes and overlap

2) Large-resolution images are imported and divided into inputs of a specific size

3) Build the network architecture shown in the preceding figure to perform network training.

4) Send the cropped image blocks to the trained network to perform forward inference and obtain the corresponding BB(weight).

5) Combine the results of multiple image blocks, a repetition rate of 15

6) Perform NMS on top of the synthesized results to suppress duplicate experiments.

B. Object Tracking

1) Challenges:

Bernardin [7] considers the audio-visual tracking of multiple persons to be a very active research field with applications in many domains. These range from video surveillance to automatic indexing and intelligent interactive environments. Especially in the last case, a robust person-tracking module can serve as a powerful building block to support other techniques, such as gesture recognizers, face or speaker identifiers, and head pose estimators.

Following the application of the mentioned detection algorithm, object confirmation and annotation were successfully executed. The subsequent phase involves the meticulous tracking of trajectories associated with individual objects.

Cannons [8] maintains the domain of multi-object target tracing and delineates two fundamental modalities: online and offline tracing. Online target tracing entails the real-time monitoring and tracking of objects, necessitating the concurrent analysis of historical and contemporaneous data streams. In contrast, offline tracing necessitates a comprehensive traversal through the entire dataset, thus demanding a priori acquaintance with the complete data inventory. Consequently, the online variant of target tracing emerges as a more intricate and resource-intensive endeavor.

Currently, prevailing research in the realm of target tracing prominently converges upon its nexus with target detection. The connotation is that the efficacy of target tracing is contingent upon the proficiency of preceding target detection operations. Notably, target tracing can be construed as an evolutionary augmentation of the fundamental task of target detection.

To expand upon this paradigm, target detection effectively demarcates bounding boxes encapsulating discrete objects in each frame of an image sequence. Subsequently, the target tracing endeavor is primed to effectuate a correlative matching process across diverse images, wherein identical objects and their respective bounding boxes are seamlessly connected. This orchestration hinges upon the discernment of unique object traits such as visual appearance and dynamic locomotion patterns. For illustrative purposes, consider an image wherein the detection algorithm proficiently identifies all objects, appending distinctive yellow-hued enclosures around them. The ensuing tracing algorithm is subsequently invoked to differentially discriminate among these diverse entities by means of variably shaded bounding boxes.

2) Related Work:

We conducted an extensive investigation into target tracking algorithms, and the research findings are illustrated in Fig.6. Several representative algorithms are highlighted, including the following.

1) RRC [9]: Breaks objects into components and tracks each component individually. Overcomes occlusion but has limitations for tracking tiny or high-density objects.

2) DeepSORT [10]: Uses Kalman filtering and Hungarian algorithm for tracking objects. Handles appearance changes and object occlusion but requires other methods for simultaneous tracking of multiple objects.

3) CenterTrack [11]: Represents objects using center points, sizes, and offsets. Effective in dense targets and target occlusion but limited in handling object deformations and appearance changes.

4) SiamRPN++ [12]: Uses IoU calculation and data augmentation to improve tracking of tiny objects. Not suitable for long-term tracing.

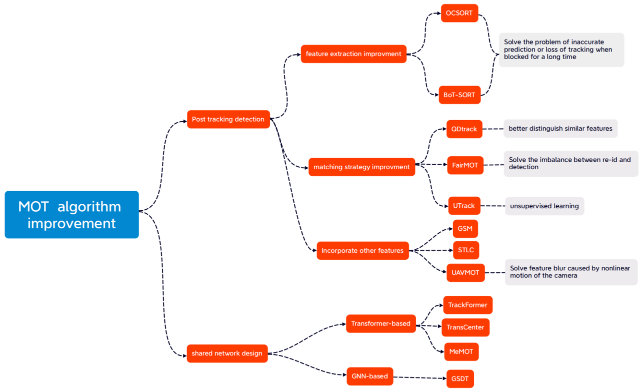

MOT algorithm improvement
Post tracking detection
feature extraction improvment
OCSORT
BoT-SORT
Solve the problem of inaccurate prediction or loss of tracking when blocked for a long time
matching strategy improvment
QDtrack
better distinguish similar features
FairMOT
Solve the imbalance between re-id and detection
UTrack
unsupervised learning
Incorporate other features
GSM
STLC
UAVMOT
Solve feature blur caused by nonlinear motion of the camera
shared network design
Transformer-based
TrackFormer
TransCenter
MeMOT
GNN-based
GSDT

Fig. 6. Related Work in Object Tracking.

3) Solution:

This methodology adheres to the tracking-by-detection paradigm and serves as a framework for multi-object tracking. In conventional multi-object tracking approaches, the establishment of target identities often relies on associating detection boxes that surpass a predefined threshold. However, this conventional method encounters substantial challenges when confronted with targets characterized by lower detection scores, notably those afflicted by occlusion.

To mitigate these challenges, a novel and versatile data association strategy, known as BYTE [13], has been introduced. BYTE departs from the traditional approach by considering the tracking of every detection box, irrespective of its detection score. By employing a principled approach, BYTE harnesses the intrinsic resemblance between low-score detection boxes and pre-existing trajectories to facilitate the recovery of authentic target identities. Moreover, it effectively eliminates spurious background detections. This innovative approach strives to enhance tracking performance, particularly in scenarios that encompass intricate attributes such as occluded targets, by comprehensively considering all available information and capitalizing on the synergy between detection and tracking mechanisms.

The most essential part of ByteTrack is data association, and we propose a simple, effective, and generic data association method, BYTE. Different from previous methods, which only keep the high-score detection boxes, we keep almost every detection box and separate them into high-score ones and low-score ones. We first associate the high-score detection boxes with the tracklets. Some tracklets get unmatched because they do not match an appropriate high score detection box, which usually happens when occlusion, motion blur, or size change occurs. We then associate the low-score detection boxes and these unmatched tracklets to recover the objects in low-score detection boxes and filter out background simultaneously. The pseudo-code of BYTE is shown in Algorithm 1.

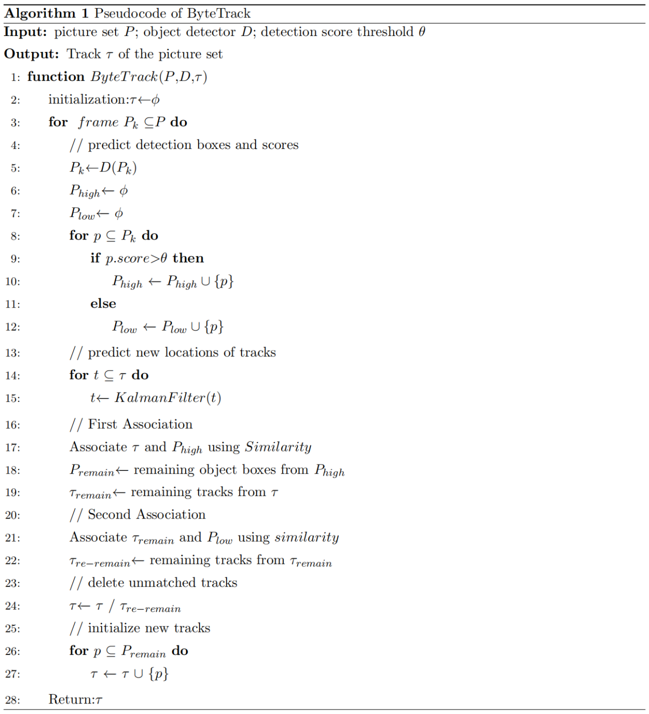
After the association, the unmatched tracks are deleted from the tracklets. We do not list the procedure of track rebirth in algorithm 1 for simplicity. It is necessary for long-range association to preserve the identity of the tracks. We put the unmatched tracks remaining after the second association into lost. For each track in lost, only when it exists for more than a certain number of frames, i.e. 30, do we delete it from the tracks. Otherwise, we keep the lost tracks in lost (line 22 in Algorithm 1). Finally, we initialize new tracks from the unmatched high-score detection boxes remaining after the first association (lines 23 to 27 in algorithm 1). The output of each individual frame comprises the bounding boxes and identities of the tracks in the current frame. Note that we do not output the boxes and identities of lost.

To put forward the state-of-the-art performance of MOT, we design a simple and strong tracker, named ByteTrack, by equipping the high-performance detector YOLOX with our association method BYTE.

C. Experiment

1) Challenges:

In the experimental phase, we encountered several significant challenges that demanded innovative solutions. One of the primary hurdles was the scarcity of suitable datasets for remote



```
Algorithm 1 Pseudocode of ByteTrack
Input: picture set $P$; object detector $D$; detection score threshold $\theta$
Output: Track $\tau$ of the picture set
function $ByteTrack(P,D,\tau)$
    initialization:$\tau \leftarrow \phi$
    for $frame\ P_k \subseteq P$ do
        // predict detection boxes and scores
        $P_k \leftarrow D(P_k)$
        $P_{high} \leftarrow \phi$
        $P_{low} \leftarrow \phi$
        for $p \subseteq P_k$ do
            if $p.score > \theta$ then
                $P_{high} \leftarrow P_{high} \cup \{p\}$
            else
                $P_{low} \leftarrow P_{low} \cup \{p\}$
        // predict new locations of tracks
        for $t \subseteq \tau$ do
            $t \leftarrow KalmanFilter(t)$
        // First Association
        Associate $\tau$ and $P_{high}$ using $Similarity$
        $P_{remain} \leftarrow$ remaining object boxes from $P_{high}$
        $\tau_{remain} \leftarrow$ remaining tracks from $\tau$
        // Second Association
        Associate $\tau_{remain}$ and $P_{low}$ using $similarity$
        $\tau_{re-remain} \leftarrow$ remaining tracks from $\tau_{remain}$
        // delete unmatched tracks
        $\tau \leftarrow \tau\ /\ \tau_{re-remain}$
        // initialize new tracks
        for $p \subseteq P_{remain}$ do
            $\tau \leftarrow \tau \cup \{p\}$
    Return:$\tau$
```

Fig. 7. Bytetrack Algorithm.

sensing image target tracking. This scarcity restricted our ability to comprehensively evaluate and validate the performance of our algorithms under diverse real-world scenarios. Additionally, determining an optimal approach to effectively showcase the results of our target tracking methodology proved to be a considerable challenge. Striking the right balance between clear visualization and technical insight posed an intriguing task, as the complexity of remote sensing data required an intuitive presentation that retained the intricacies of our tracking algorithm's outcomes. Overcoming these obstacles required a multidisciplinary approach, leveraging expertise in both remote sensing data acquisition and algorithmic visualization techniques to ensure the accurate representation and meaningful interpretation of our algorithm's tracking results.

2) *Solution:*

To solve the challenge of the dataset, we procured our dataset from a reputable academic challenge known as the ”1st Challenge on Moving Object Detection and Tracking in Satellite.” This comprehensive dataset was captured by the Jilin-1 satellite constellation from various positions along its orbit. The specific small object we are interested in detecting and tracking within these images is cars. Our training dataset comprises a total of 800 images, each measuring 1024 by 1024 pixels. Notably, the small cars we aim to identify are roughly 5 by 5 pixels in size. Further, in order to enhance result clarity and facilitate straightforward model usage, we have developed a graphical user interface application based on PyQt5. The visual interface is shown in Fig.8. The users can press the import button to import a photo or a video, and then press the predict button to show the object tracking result.

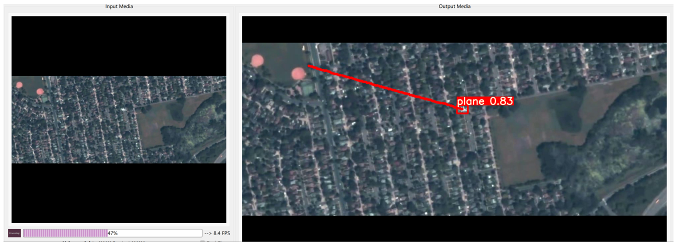


Fig. 8. The Visual Interface.

VI. Contributions and Limitations

In summary, this project aims to combine existing efforts from both domestic and international researchers to achieve the detection and tracking of small targets in remote sensing images. Building upon prior work, we endeavor to introduce innovations that enhance the performance of small target detection and tracking in remote sensing imagery. The contributions of our project are as follows. Firstly, we conducted deep research into deep learning-based object detection and tracking for remote sensing images. Secondly, we used improved YOLOv5 and Bytetrack to achieve detection and tracking of small targets in remote sensing images. Thirdly we designed a visual interface to display our object detection and tracking results. Due to time constraints, there are many tasks that we have not been able to complete. Our results lack comparisons with other algorithms, and we still need to improve our algorithm. Based on the guidance of our professor, our future work will unfold in three aspects: firstly, incorporating super-resolution before target detection to enhance target feature intensity; secondly, leveraging techniques such as Spatial Attention Module (SAM) [14] or other semantic segmentation networks to elevate network performance; and thirdly, synergizing traditional machine learning approaches with deep learning methods. We will strive to make breakthroughs in these three directions to further enhance the accuracy of target detection and tracking.

References

[1] Li, K., Wan, G., Cheng, G., Meng, L., & Han, J. (2020). Object detection in optical remote sensing images: A survey and a new benchmark. ISPRS journal of photogrammetry and remote sensing, 159, 296-307.

[2] Toth, C., & Jóźków, G. (2016). Remote sensing platforms and sensors: A survey. ISPRS Journal of Photogrammetry and Remote Sensing, 115, 22-36.

[3] Chi, M., Plaza, A., Benediktsson, J. A., Sun, Z., Shen, J., & Zhu, Y. (2016). Big data for remote sensing: Challenges and opportunities. Proceedings of the IEEE, 104(11), 2207-2219.

[4] Song, J., Gao, S., Zhu, Y., & Ma, C. (2019). A survey of remote sensing image classification based on CNNs. Big earth data, 3(3), 232-254.

[5] Fatima, S. A., Kumar, A., Pratap, A., & Raoof, S. S. (2020, January). Object recognition and detection in remote sensing images: a comparative study. In 2020 International Conference on Artificial Intelligence and Signal Processing (AISP) (pp. 1-5). IEEE.

[6] Mao, J., Zhang, X., Ji, Y., Zhang, Z., & Guo, Z. (2021, June). Improved high precision aircraft target detection method of yolt. In Journal of Physics: Conference Series (Vol. 1955, No. 1, p. 012028). IOP Publishing.

[7] Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: the clear mot metrics. EURASIP Journal on Image and Video Processing, 2008, 1-10.

[8] Cannons, K. (2008). A review of visual tracking. Dept. Comput. Sci. Eng., York Univ., Toronto, Canada, Tech. Rep. CSE-2008-07, 242.

[9] Lukežič, A., Zajc, L. Č., & Kristan, M. (2017). Deformable parts correlation filters for robust visual tracking. IEEE transactions on cybernetics, 48(6), 1849-1861.

[10] Veeramani, B., Raymond, J. W., & Chanda, P. (2018). DeepSort: deep convolutional networks for sorting haploid maize seeds. BMC bioinformatics, 19, 1-9.

[11] Zhang, H., Chen, Y., Wang, J., & Yang, S. (2015). Cycle-based optimal NOx emission control of selective catalytic reduction systems with dynamic programming algorithm. Fuel, 141, 200-206.

[12] Zhou, X., Koltun, V., & Krähenbühl, P. (2020, August). Tracking objects as points. In European conference on computer vision (pp. 474-490). Cham: Springer International Publishing.

[13] Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., ... & Wang, X. (2022, October). Bytetrack: Multi-object tracking by associating every detection box. In European Conference on Computer Vision (pp. 1-21). Cham: Springer Nature Switzerland.

[14] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. arXiv preprint arXiv:2304.02643.