Computers and Artificial Intelligence；Vol.3, No.3; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n3.6673
Published by: Zeus Press

A Real-Time Multi-Target Capture Framework Based on a Single PTZ Camera

Yi Tian[1,3] and Yang Yang[2,3]*

[1]*School of Physics and Electronic Information, Yunnan Normal University, Kunming, Yunnan, China*

[2]*School of Information Science and Technology, Yunnan Normal University, Kunming, Yunnan, China*

[3]*Laboratory of Pattern Recognition and Artificial Intelligence, Yunnan Normal University, Kunming, Yunnan, China*

**Corresponding author: Yang Yang.*

Abstract

To address the problems of insufficient target details and low multi-target capture efficiency in large-scale surveillance scenes, this paper proposes a real-time multi-target capture framework based on a single PTZ (Pan-Tilt-Zoom) camera. The proposed method integrates object detection, PTZ parameter estimation, and path scheduling to achieve rapid sequential capture of multiple targets. First, an object detection algorithm is used to locate targets in the scene. Then, the corresponding PTZ control parameters are estimated according to the pixel coordinates of each target, and feasible capture regions are constructed. Finally, a scheduling strategy combining greedy path planning and local optimization is adopted to generate a low-cost capture sequence for rapid PTZ switching among multiple targets. Experimental results demonstrate that the proposed framework can effectively reduce PTZ motion time and improve multi-target capture efficiency while maintaining stable operation in real-world environments. The proposed framework requires no additional hardware and has good practical deployment value for intelligent surveillance applications.

Keywords

PTZ camera, multi-target capture, path planning, object detection, intelligent surveillance

1. Introduction

Rapid multi-target capture in large-scale scenes with a single PTZ camera is an important and practically motivated research direction. From a societal perspective, although PTZ cameras have been widely deployed, in practice they typically only perform basic functions such as continuous recording and post-event playback. They lack the capability for rapid switching among multiple targets and high-quality snapshot capture. If existing PTZ cameras can be enabled to perform rapid multi-target capture without adding new hardware or infrastructure, the efficiency and responsiveness of public safety and urban governance will be significantly improved. Rapid multi-target capture with a single PTZ camera is also essential in many specialized applications, including military reconnaissance and border patrol, ecological and environmental monitoring, and abnormal event response in industrial parks. Deployment cost constraints often make multi-camera

systems or additional hardware impractical. Therefore, rapid multi-target capture using a single PTZ camera remains a research topic of strong application significance.

However, there remains a significant gap between coarse-grained target detection and fine-grained recognition in large-scale scenarios. Current research on multi-target capture based on PTZ cameras can be broadly categorized into three strategies:

(1) Multi-camera collaborative systems integrate fixed cameras and PTZ cameras with complementary functions, aiming to achieve continuous target visibility and stable spatiotemporal coverage through functional specialization and coordinated control. A typical configuration employs a fixed-PTZ hybrid architecture: static wide-angle cameras provide global detection and guidance, while PTZ units execute high-resolution directional capture [1-6].

(2) Hardware-based PTZ responsiveness enhancement seeks to reduce motion latency and boost responsiveness through hardware innovation. For example, galvanomirror-based ultrafast pan--tilt cameras enable millisecond-level viewpoint switching and time-division multiplex a single telephoto view to simulate multiple virtual cameras, thereby supporting high-frequency multi-target capture [7, 8]. High-speed imaging modules that combine high-frame-rate CMOS sensors with long focal-length lenses can achieve over 500 fps for real-time zoom tracking [9], while fast focus stacking extends the depth-of-field during rapid zoom transitions [10]. In addition, stereo vision modules can further reduce localization latency, enabling fast 3D positioning and near-simultaneous high-resolution acquisition of multiple targets [11].

(3) Viewpoint planning and scheduling with a single PTZ camera focuses on coordinating pan, tilt, and zoom actions to achieve efficient and high-quality capture under limited sensing resources. Information-theoretic methods often adopt greedy scheduling driven by immediate information gain maximization to determine step-by-step PTZ viewpoint adjustments [12, 13]. For face association tasks that require fine-grained recognition, heuristic scoring-based frameworks repeatedly choose the highest-priority (or highest-scoring) target for zoom-in capture, then return to wide-angle monitoring, forming alternating cycles between global observation and focused identity resolution [14, 15]. More recently, deep reinforcement learning has been introduced to optimize viewpoint adjustments, adaptively prioritizing low-confidence targets via proactive zoom operations, and thereby improving the imaging quality of small or distant objects [16].

Therefore, developing a real-time multi-target capture method based on a single conventional PTZ camera has significant practical value. In this paper, we propose a real-time multi-target capture framework that combines object detection, PTZ parameter estimation, and path optimization to achieve efficient sequential capture of multiple targets using only a single PTZ camera.

The main contributions of this paper are summarized as follows:

1) A real-time multi-target capture framework based on a single PTZ camera is proposed.
2) A mapping method from image pixel coordinates to PTZ control parameters is designed.
3) A path planning strategy combining greedy search and local optimization is developed.
4) Real-world experiments are conducted to verify the effectiveness of the proposed framework.

2. Related Work

2.1 Multi-Camera Collaborative Surveillance

Multi-camera collaborative surveillance systems are one of the most widely used solutions for large-scale target monitoring. These methods usually combine fixed wide-angle cameras and PTZ cameras to achieve continuous target observation and high-resolution capture. Fixed cameras are responsible for global target detection and scene understanding, while PTZ cameras perform local zoom-in capture and detailed observation. Early studies mainly focused on target association and cross-view tracking among multiple cameras. Later works introduced collaborative scheduling and distributed control strategies to improve target coverage and tracking stability [18, 19]. Although multi-camera systems can significantly improve surveillance capability, they generally require complex calibration, additional communication infrastructure, and higher deployment costs, which limits their applicability in resource-constrained environments.

2.2 Hardware-Enhanced PTZ Systems

Another important research direction focuses on improving PTZ responsiveness through specialized hardware designs. Some studies adopted galvanometer-based ultra-fast steering systems to achieve millisecond-level viewpoint switching and high-frequency target observation. Other methods combined high-speed CMOS sensors with long-focus optical systems to support rapid zoom tracking and real-time image acquisition. Fast focus stacking techniques were also introduced to improve image sharpness during rapid zoom transitions. In addition, stereo vision systems and depth-assisted PTZ platforms have been proposed to reduce target localization latency and improve three-dimensional positioning accuracy. Although these methods can achieve excellent capture performance, they usually rely on expensive customized hardware devices and are difficult to deploy on conventional PTZ surveillance systems.

2.3 Single PTZ Camera Scheduling

Viewpoint scheduling and path optimization for a single PTZ camera have also attracted increasing attention in recent years. Existing approaches mainly focus on determining efficient pan, tilt, and zoom adjustment sequences under limited sensing resources. Information-theoretic methods generally adopt greedy strategies to maximize immediate observation gain. Heuristic scheduling frameworks prioritize targets according to confidence scores or task importance and sequentially perform zoom-in capture. More recently, reinforcement learning methods have been introduced to learn adaptive viewpoint adjustment policies for dynamic surveillance environments. However, many existing methods simplify PTZ motion modeling and lack consideration of realistic camera actuation costs, which limits their practical performance in real-world applications.

Compared with the above approaches, the proposed framework focuses on achieving efficient multi-target capture using only a single conventional PTZ camera without requiring additional hardware or multi-camera collaboration. By integrating object detection, PTZ parameter estimation, and path optimization, the proposed method provides a practical and deployable solution for real-time multi-target surveillance.

3. Method

3.1 System Framework

The proposed framework mainly consists of four modules: object detection, PTZ parameter estimation, path planning, and PTZ control. The overall workflow is illustrated as a closed-loop capture process that enables a single PTZ camera to perform efficient multi-target observation and sequential high-resolution capture in large-scale scenes.

First, the PTZ camera operates at a relatively low zoom level to obtain a wide-angle surveillance image covering a large monitoring area. The captured video stream is continuously transmitted to the object detection module, where a deep-learning-based object detector is employed to identify targets in the scene and generate corresponding bounding box information. Each detected target is represented by its pixel coordinates and image position, which provide the basis for subsequent PTZ parameter estimation.

After target detection, the PTZ parameter estimation module converts the image-space target coordinates into the corresponding PTZ control parameters. Specifically, the center position of each target in the image is used to estimate the required pan and tilt rotation angles, while the target size and image quality requirements are utilized to estimate the appropriate zoom ratio. Based on these parameters, a feasible capture region is constructed for each target in the PTZ control space, representing all valid camera states capable of capturing the target clearly while maintaining the target near the center of the image.

Once the PTZ parameters of all targets are obtained, the path planning module determines an efficient capture sequence for the camera. Since PTZ cameras require physical motion when switching between different viewpoints, different target visiting orders may lead to significantly different capture times. To reduce unnecessary camera movement and improve capture efficiency, the framework adopts a scheduling strategy combining greedy initialization and local optimization. The generated capture sequence minimizes the overall PTZ motion cost, including pan rotation, tilt adjustment, and zoom variation between consecutive targets.

Finally, the PTZ control module executes the optimized capture sequence and drives the camera to sequentially perform target capture. The control commands are transmitted to the PTZ device through the camera SDK or network control interface, enabling automatic pan, tilt, and zoom adjustment during the capture process. After the camera reaches the desired viewpoint, high-resolution target snapshots are obtained and stored for subsequent analysis or monitoring applications.

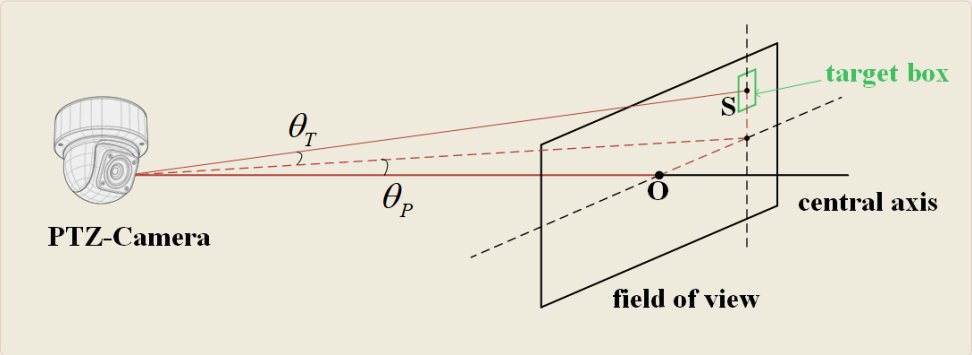
To improve the real-time performance and responsiveness of the system, the proposed framework adopts a multi-threaded architecture. Video acquisition, object detection, path planning, and PTZ control are executed independently in parallel threads, preventing computational delays in one module from blocking the entire system. In particular, asynchronous detection and non-blocking PTZ control significantly improve system stability under dynamic surveillance conditions, allowing the framework to maintain continuous operation in real-world multi-target capture scenarios.

3.2 PTZ Parameter Estimation

3.2.1 Angle Inversion Module

To achieve reliable pan-tilt estimation based on image spatial coordinates, we reconstruct and optimize the inverse mapping method, deriving an explicit expression based on the geometric principles of PTZ imaging. The final angle inversion model calculates the required tilt and translation angles as follows:

Figure 1: Schematic diagram of the principle of calculating P, T values



Based on the camera's sensor specifications and display resolution, the effective horizontal and vertical dimensions of the imaging area, denoted as H_{size} and V_{size} (in millimeters), are obtained. Then, given a target's 2D pixel coordinates(X,Y), a linear mapping is applied to project these values onto the physical image sensor plane as (x,y) using the following equations:

$$x=\frac{X\cdot H_{size}}{H_{pix\_total}},\ y=\frac{Y\cdot V_{size}}{V_{pix\_total}} \tag{1}$$

where H_{pix_total} and V_{pix_total} are the total number of pixels in the horizontal and vertical directions, respectively.

(i) shows the field of view angles under two zoom conditions; (ii) and (iii) illustrate the maximum and minimum required field

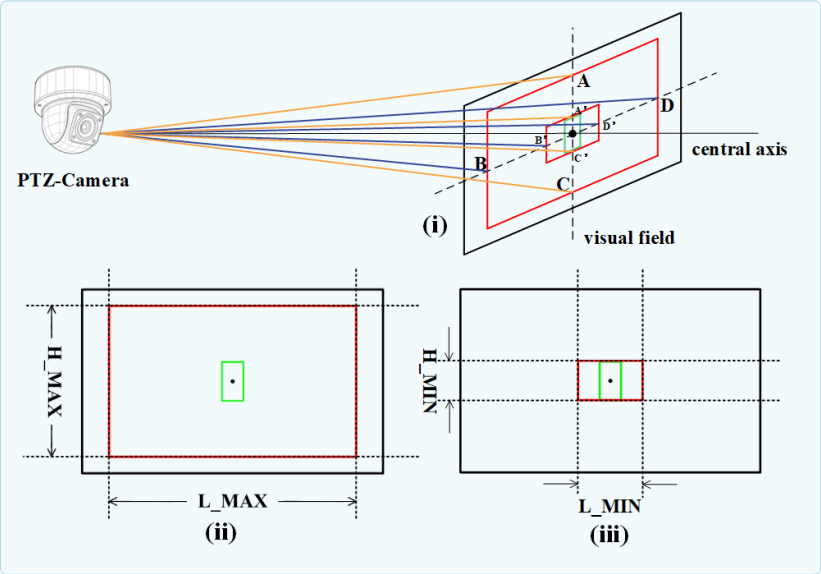
The current camera status can be obtained via the camera SDK: focal length f(in mm), pan angle β, and tilt angle α. Using geometric relations, the angular deviations between the target's center point and the optical axis are calculated in both horizontal β_c and vertical α_c directions:

$$\theta_T=\arcsin\left(\frac{\sqrt{(y-y_0)^2+f^2}}{\sqrt{(x-x_0)^2+(y-y_0)^2+f^2}}\times\sin\left(\arctan\frac{y-y_0}{f}+\alpha\right)\right)-\alpha \tag{2}$$

$$\theta_P=\arcsin\left(\sin\left(\arctan\frac{|x-x_0|}{\sqrt{(y-y_0)^2+f^2}}\right)/\cos\theta_T\right) \tag{3}$$

where (x_0,y_0) is the center of the image sensor. The calculated angles are shown in Fig.1. After this transformation, the required rotation angle for the PTZ camera to display the target at the screen center can be derived.

Figure 2



3.2.2 Zoom Estimation Module

Zooming in on a scene can essentially be understood as spatially cropping a specific area from the original panoramic image (1× zoom), thereby achieving magnification of the local field of view. This process not only reduces the observable area but also enhances the rendering of details.

By fixing the pan-tilt position, we utilized the official camera SDK to control the PTZ camera to move to a specified zoom level and capture the corresponding field of view (FOV) for each zoom setting. This enabled us to establish the relationship between zoom values and field of view, and subsequently model it through data-driven function fitting.

$$FOVH_{need} = 2 \cdot \arctan\left(\frac{fov_L}{2560} \cdot \tan\left(\frac{FOV_x}{2} \cdot \frac{\pi}{180}\right)\right) \cdot \frac{180}{\pi} \tag{4}$$

$$FOVV_{need} = 2 \cdot \arctan\left(\frac{fov_H}{1440} \cdot \tan\left(\frac{FOV_y}{2} \cdot \frac{\pi}{180}\right)\right) \cdot \frac{180}{\pi} \tag{5}$$

$$z_1 = \frac{k_{H2}}{\ln\left(\frac{FOVH_{need}}{k_{H1}}\right)} - k_{H3} \tag{6}$$

$$z_2 = \frac{k_{V2}}{\ln\left(\frac{FOVV_{need}}{k_{V1}}\right)} - k_{V3} \tag{7}$$

$$z = \min(z_1, z_2) \tag{8}$$

where FOV_x and FOV_y are the horizontal and vertical field-of-view angles at 1× focal length, respectively. $FOVH_{need}$ and $FOVV_{need}$ denote the required horizontal and vertical field-of-view angles. The coefficientsk_{H1}, k_{H2}, k_{H3}, k_{V1}, k_{V2}, and k_{V3} vary across camera models and are obtained by calibration (curve fitting). Where z_1 and z_2 are the zoom ratio calculated from the horizontal field of view and the zoom ratio calculated from the vertical field of view, respectively.

In target recognition, detection performance is strongly affected by the target's image proportion. In COCO, objects larger than 96×96 pixels are regarded as "large" [17]. Under a typical resolution (e.g.,640×480) this corresponds to an area ratio of roughly 3%. In our 2560×1440 images, the same absolute size occupies only about 0.2% of the image area and is therefore relatively "tiny." To ensure sufficient pixel evidence for reliable detection and recognition (e.g., for YOLOv8), we set a stricter visibility requirement: the target should cover at least 5% of the image area. This ratio yields the minimum observation scale (Z-value) needed to satisfy the visibility constraints. The maximum zoom is jointly limited by Fig.2(i) the camera's optical zoom capability and Fig.2 (ii) the requirement that the zoomed target bounding box remains fully inside the image. Equivalently, zoom can be interpreted as cropping a subwindow from the 1× view: a smaller subwindow implies a larger zoom factor and a narrower field of view. As illustrated in Fig.3 (ii), once the subwindow is chosen to meet the5% area requirement, the corresponding horizontal and vertical FOV angles can be computed by Eq. (3.4) and (3.5), which determines the required Z value. Fig.2(iii) shows the limiting case of the maximum magnification, i.e., the largest Z for which the target is still fully visible.

This method first calculates the required field of view based on the scaling ratio of images within the standard resolution. Subsequently, it derives the corresponding zoom value through a log model fitted experimentally. Finally, it averages the horizontal and vertical field-of-view angle results and converts the outcome into the format required by the camera SDK.

3.3 Multi-Target Capture Scheduling

When switching between different targets, the PTZ camera must perform pan rotation, tilt rotation, and zoom adjustment. Different capture orders therefore lead to different total motion costs.

To reduce the overall capture time, this paper adopts a scheduling strategy that combines greedy search and local optimization.

First, starting from the current PTZ state, the system selects the next target with the minimum motion cost to generate an initial capture sequence. Then, a 2-opt local optimization method is applied to further refine the target visiting order and reduce the total scheduling cost.

The motion cost calculation mainly considers:

- Pan rotation time;
- Tilt rotation time;
- Zoom adjustment time.

Through iterative optimization of the target visiting order, the proposed method achieves lower overall capture time for multi-target scheduling.

4. Experiments and Results

To evaluate the effectiveness of the proposed framework, experiments were conducted on a real PTZ camera platform. The experimental scenarios include outdoor pedestrian scenes and open-area multi-target capture environments.

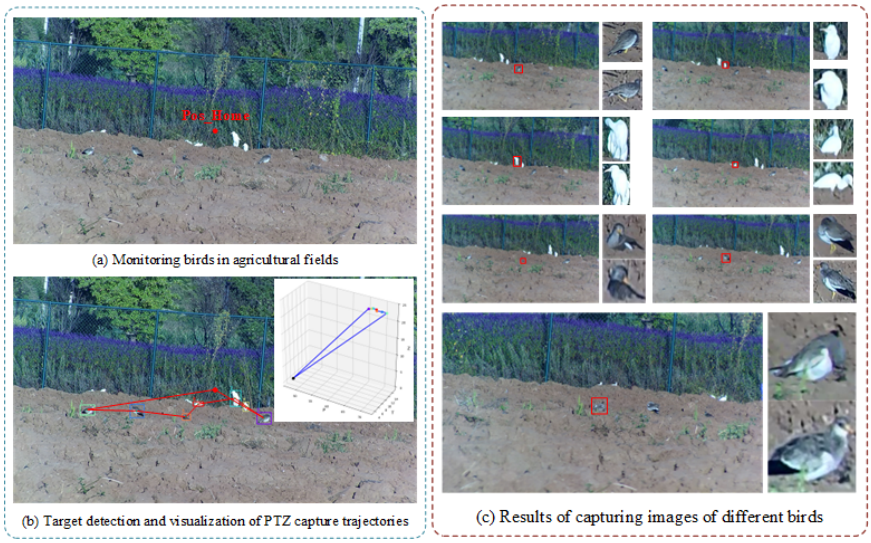
During the experiments, the system successfully detected multiple targets and automatically generated optimized capture sequences. Compared with traditional sequential capture methods based directly on detection order, the proposed framework significantly reduced PTZ motion time and improved capture efficiency.

Experimental results demonstrate that:

1) The system can stably perform continuous multi-target capture;
2) The proposed scheduling strategy effectively reduces overall motion time;
3) The framework achieves good real-time performance and practical deployment capability.

In addition, the proposed method requires no additional hardware and can be deployed using only a conventional PTZ camera, making it suitable for practical intelligent surveillance applications.

Figure 3: Surveillance of birds on an embankment



The camera is positioned approximately 100m from the scene, with the lens set to 5× zoom. The resulting image clearly shows details such as feather color, beak color and size, and overall body shape.

5. Conclusion

This paper proposes a real-time multi-target capture framework based on a single PTZ camera. The proposed method integrates object detection, PTZ parameter estimation, and path scheduling to achieve efficient sequential capture of multiple targets. Experimental results demonstrate that the framework can effectively improve PTZ capture efficiency while maintaining stable real-world performance. Future work will focus on dynamic target prediction and adaptive scheduling in more complex surveillance environments.

References

[1] Del Bimbo A, Pernici F. Towards on-line saccade planning for high-resolution image sensing[J]. Pattern Recognition Letters, 2006, 27(15): 1826-1834.

[2] Bellotto N, Sommerlade E, Benfold B, et al. A distributed camera system for multi-resolution surveillance[C]//2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC). IEEE, 2009: 1-8.

[3] Liu Y, Lai S, Zuo C, et al. A Master-Slave Surveillance System to Acquire Panoramic and Multiscale Videos[J]. The Scientific World Journal, 2014, 2014(1): 491549.

[4] Chen C H, Yao Y, Page D, et al. Heterogeneous fusion of omnidirectional and PTZ cameras for multiple object tracking[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2008, 18(8): 1052-1063.

[5] Natarajan P, Hoang T N, Low K H, et al. Decision-theoretic approach to maximizing observation of multiple targets in multi-camera surveillance[C]//AAMAS. 2012: 155-162.

[6] Yi J, Acer U G, Kawsar F, et al. Argus: Enabling cross-camera collaboration for video analytics on distributed smart cameras[J]. IEEE Transactions on Mobile Computing, 2024, 24(1): 117-134.

[7] Hu S, Shimasaki K, Jiang M, et al. A dual-camera-based ultrafast tracking system for simultaneous multi-target zooming[C]//2019 IEEE International Conference on Robotics and Biomimetics (ROBIO). IEEE, 2019: 521-526.

[8] Hu S, Shimasaki K, Jiang M, et al. A simultaneous multi-object zooming system using an ultrafast pan-tilt camera[J]. IEEE Sensors Journal, 2021, 21(7): 9436-9448.

[9] Li Q, Hu S, Shimasaki K, et al. An active multi-object ultrafast tracking system with CNN-based hybrid object detection[J]. Sensors, 2023, 23(8): 4150.

[10] Zhang T, Li Z, Wang Q, et al. Dof-extended zoomed-in monitoring system with high-frame-rate focus stacking and high-speed pan-tilt adjustment[J]. IEEE Sensors Journal, 2024, 24(5): 6765-6776.

[11] Li Q, Hu S, Shimasaki K, et al. An Ultrafast Multi-object Zooming System Based on Low-latency Stereo Correspondence[C]//2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2024: 11552-11557.

[12] Sommerlade E, Reid I. Information-theoretic active scene exploration[C]//2008 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2008: 1-7.

[13] Salvagnini P, Pernici F, Cristani M, et al. Information theoretic sensor management for multi-target tracking with a single pan-tilt-zoom camera[C]//IEEE Winter Conference on Applications of Computer Vision. IEEE, 2014: 893-900.

[14] Cai Y, Medioni G, Dinh T B. Towards a practical PTZ face detection and tracking system[C]//2013 IEEE Workshop on Applications of Computer Vision (WACV). IEEE, 2013: 31-38.

[15] Melman S, Moses Y, Medioni G, et al. The multi-strand graph for a PTZ tracker[C]//2015 12th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). IEEE, 2015: 1-6.

[16] Fang H, Liu H, Wen J, et al. Automatic visual enhancement of PTZ camera based on reinforcement learning[J]. Neurocomputing, 2025, 626: 129531.

[17] Lin T Y, Maire M, Belongie S, et al. Microsoft coco: Common objects in context[C]//European conference on computer vision. Cham: Springer International Publishing, 2014: 740-755.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights