ORIGINAL RESEARCH
published: 20 May 2022
doi: 10.3389/fbioe.2022.905583

Biomimetic Vision for Zoom Object Detection Based on Improved Vertical Grid Number YOLO Algorithm

Xinyi Shen[1], *Guolong Shi*[1,2]*, *Huan Ren*[1] *and Wu Zhang*[1]

[1]School of Information and Computer, Anhui Agricultural University, Hefei, China, [2]School of Electrical Engineering and Automation, Wuhan University, Wuhan, China

With the development of bionic computer vision for images processing, researchers have easily obtained high-resolution zoom sensing images. The development of drones equipped with high-definition cameras has greatly increased the sample size and image segmentation and target detection are important links during the process of image information. As biomimetic remote sensing images are usually prone to blur distortion and distortion in the imaging, transmission and processing stages, this paper improves the vertical grid number of the YOLO algorithm. Firstly, the light and shade of a high-resolution zoom sensing image were abstracted, and the grey-level cooccurrence matrix extracted feature parameters to quantitatively describe the texture characteristics of the zoom sensing image. The Simple Linear Iterative Clustering (SLIC) superpixel segmentation method was used to achieve the segmentation of light/dark scenes, and the saliency area was obtained. Secondly, a high-resolution zoom sensing image model for segmenting light and dark scenes was established to made the dataset meet the recognition standard. Due to the refraction of the light passing through the lens and other factors, the difference of the contour boundary light and dark value between the target pixel and the background pixel would make it difficult to detect the target, and the pixels of the main part of the separated image would be sharper for edge detection. Thirdly, a YOLO algorithm with an improved vertical grid number was proposed to detect the target in real time on the processed superpixel image array. The adjusted aspect ratio of the target in the remote sensing image modified the number of vertical grids in the YOLO network structure by using 20 convolutional layers and five maximum aggregation layers, which was more accurately adapted to "short and coarse" of the identified object in the information density. Finally, through comparison with the improved algorithm and other mainstream algorithms in different environments, the test results on the aid dataset showed that in the target detection of high spatial resolution zoom sensing images, the algorithm in this paper showed higher accuracy than the YOLO algorithm and had real-time performance and detection accuracy.

Keywords: bionic vision, zoom target detection, deep learning, image segmentation, simple linear iterative clustering, light/dark co-occurrence scene

OPEN ACCESS

Edited by:
Gongfa Li,
Wuhan University of Science and Technology, China

Reviewed by:
Huawei Liu,
North China Electric Power University, China
Baohua Zhang,
Nanjing Agricultural University, China

***Correspondence:**
Guolong Shi
shigl@ahau.edu.cn

Specialty section:
This article was submitted to Bionics and Biomimetics, a section of the journal Frontiers in Bioengineering and Biotechnology

Received: *27 March 2022*
Accepted: *02 May 2022*
Published: *20 May 2022*

Citation:
Shen X, Shi G, Ren H and Zhang W (2022) Biomimetic Vision for Zoom Object Detection Based on Improved Vertical Grid Number YOLO Algorithm. Front. Bioeng. Biotechnol. 10:905583. doi: 10.3389/fbioe.2022.905583

INTRODUCTION

Biomimetic remote sensing is an interdisciplinary discipline that covers a variety of technical disciplines, including computer technology, sensor technology, image processing technology and other technologies (Duan et al., 2021; Chen T. et al., 2021). Biomimetic remote sensing technology uses the electromagnetic wave reflection and radiation information of the target to capture and image the electromagnetic wave reflected and emitted by the electromagnetic wave sensor installed on the spacecraft, such as satellites and hot air balloons. It has become a comprehensive new technology to process and analyse the characteristic information of specific targets. An important application field of remote sensing technology is land resource monitoring, such as land use information statistics, land cover dynamic monitoring and data updates (Lee and Ke, 2018; Giri et al., 2019; Islam et al., 2021). With the development of spatial information science, remote sensing images are widely used and provide researchers with reliable data sources. The classification and location of specific targets in remote sensing images can be obtained through target recognition technology, which can be applied to species identification and classification, crop area estimation and monitoring, crop nutrient and water status monitoring and other fields. Therefore, the detection and recognition of remote sensing targets has important research value (Shadrin et al., 2020; Yu et al., 2021; Hu et al., 2015).

Currently, geographic information systems (GISs) have been widely used in many fields. However, there is a bottleneck problem in GIS applications, that is, how to quickly extract target information and update GIS data. In the field of photogrammetry and remote sensing, target extraction and recognition are technical hot spots that are urgently needed in production applications but still far from production applications and are one of the focuses of remote sensing (Li and Liu 2020). Therefore, object extraction of images is of great significance. The objects detected by remote sensing are divided into three categories: point targets, linear targets (such as roads, rivers, etc.) and planar targets (such as buildings, etc.) (Abdollahi et al., 2020; Huang et al., 2022). Rich texture information and spatial information are important characteristics of high spatial resolution zoom sensing images that distinguish them from other images (Huang et al., 2021). High-precision classification of high-spatial-resolution images plays an important role in agriculture, urban planning, environmental monitoring and other fields. How to extract the target effectively is always a difficult problem. At present, the most commonly used target detection methods are mainly target motion detection. The literature (Mahmoodi and Salajeghe, 2019; Zhang et al., 2019; Sun et al., 2020) has proposed the background difference method, interframe difference method, optical flow method, etc., but only moving targets can be detected, while stationary targets or slow-moving targets cannot be effectively detected, and targets cannot be accurately classified (Jiang et al., 2021). Thanks to deep learning, mainly convolutional neural networks and candidate region algorithms (Haut et al., 2018; Wang and Dou et al., 2019), a huge breakthrough in target detection has been made. The literature (Jin et al., 2020; Peng et al., 2020; Khan et al., 2021) defines RCNN, Fast RCNN, Faster-RCNN and YOLO, among which YOLO is a brand-new target detection method that integrates target determination and target recognition. End-to-end detection not only achieves fast detection but also achieves better performance (Sadykova et al., 2020; Liu X. et al., 2021).

To accurately detect targets, this paper first abstracted the light and shade of high-resolution zoom sensing images and adopts SLIC superpixel segmentation. Second, a YOLO algorithm with an improved number of vertical grids was proposed to detect targets in real time on the processed superpixel image array. YOLO grid framework could divide the image into different regions, so as to predict the boundary box and probability of each region. These boundary boxes would be weighted by the predicted probability. After adjusting the vertical mesh according to the structure of the curved zoom microlens array, the prediction accuracy of the algorithm would increase and the speed would also increase. The comparison with the improved algorithm in different environments showed that the proposed method had better accuracy. At the same time, compared with other mainstream target algorithms, it was verified that the improved algorithm in this paper was suitable for scenes requiring both speed and precision. The increasing resolution of remote sensing images had become a reality for realizing agricultural work from relatively macro large-scale species identification and monitoring to individual plant species identification and monitoring of the drop changes in crop nutrition, diseases and insect pests. The improvement of image resolution would improve the precision of future agricultural work.

RELATED WORK

In computer vision, images with medium and high accuracy have become the data types commonly used by researchers, especially in species identification and classification in agriculture, such as agricultural vegetation classification, land use classification, crop classification, tree species identification, etc., (Roslim et al., 2021; Chen Z. et al., 2021; Li et al., 2021; Yan et al., 2021). To solve the common problems of zoom sensing image segmentation algorithms, such as poor robustness, easy loss of edge information and narrow scope of application, the core task of zoom sensing image target detection is to judge whether there is a target in zoom sensing images and to detect, segment, extract and classify it.

Edges contain the most concentrated rich local image information, and the Gaussian filter used for smoothing also blurs the edges of the image. In order to more accurately examine the edges in image segmentation, Liu et al. (2019) chose guaranteed edge guide filtering features and used Canny operator and wave decomposition. Each band edge detection processing was conducted, and then the edges would be integrated into a result image. Compared with the traditional edge detection operator, the segmentation result is smoother, but for remote sensing images with complex background and changeable environment, the model generalization performance is poor and the detection effect is not good. Li et al. (2019) proposed the aircraft target detection model DC-DNN of remote sensing images based on a deep neural network. This model relied on a small number of image-level labels and eliminates overlapping frames and false detection frames through a detection frame suppression algorithm to complete pixel-level target

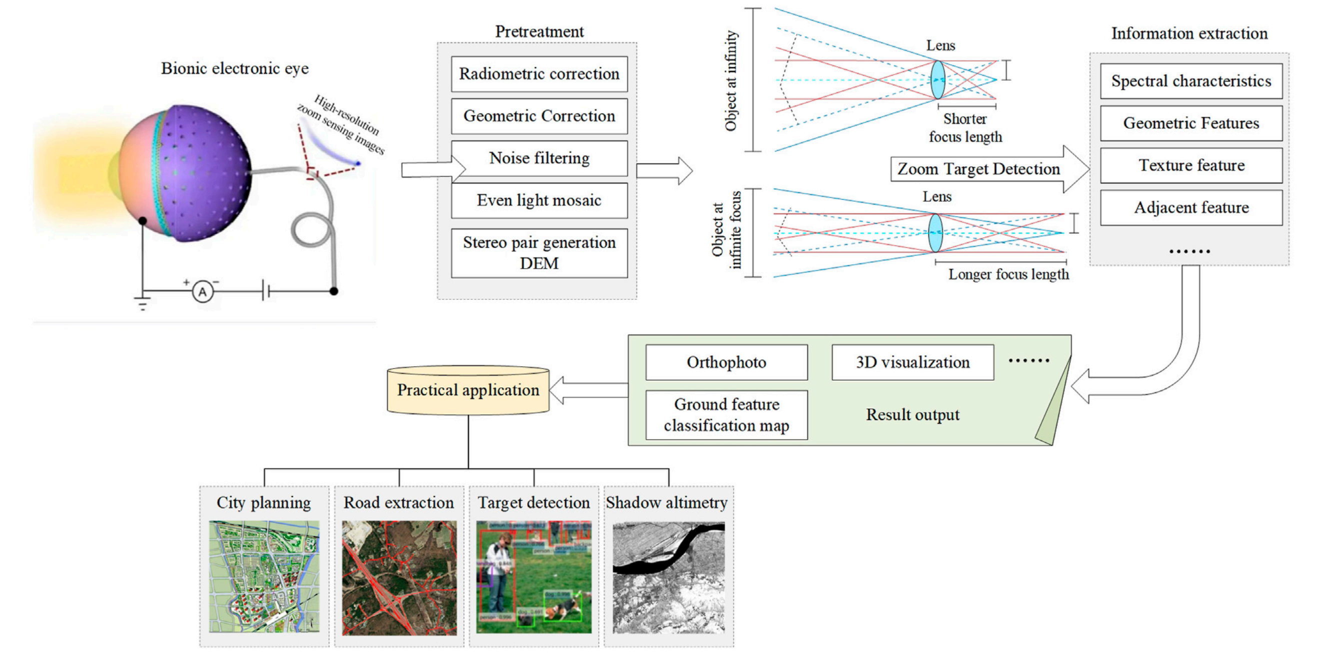

FIGURE 1 | High-resolution zoom image processing and application.

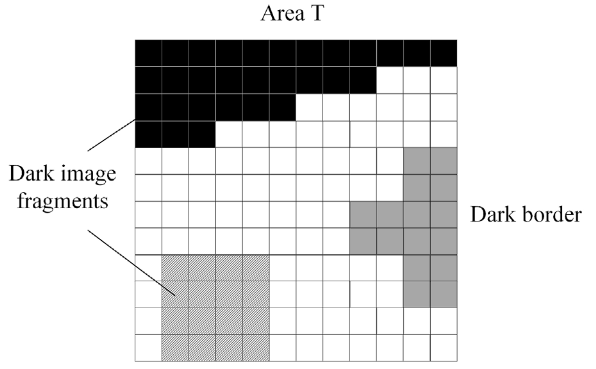

FIGURE 2 | Truncated median filter model.

segmentation and detection of zoom sensing images. The average pixel accuracy of the DC-DNN supervised FCN depth model in the three datasets was 81.47%, and the detection accuracy of aircraft targets in remote sensing images was 95.78%. Li et al. (2019) improved the Faster-RCNN network by adding the attention mechanism module into the feature extraction network to obtain more information about the target to be paid attention to and suppress other useless information to adapt to the problem of complex backgrounds and small targets caused by the large field of vision of remote sensing images. The result was 12.2% higher than that of the original Faster-RCNN. Alganci et al. (2020) compared and evaluated the performance of target detection algorithm based on deep learning convolutional neural network in aircraft target detection task in remote sensing image. Fast RCNN achieved the highest accuracy, but the detection speed was too slow to meet the needs of real-time detection. SSD algorithm had the lowest detection performance, and YOLOv3 achieved a balance between accuracy and detection speed. Hou and Jiang, (2021) improved YOLOv4's trunk feature extraction network by using dense link network multiplexing features, strengthened the detection ability of small targets, and obtained a target detection algorithm more suitable for detecting aircraft in remote sensing images, which effectively optimized the problem of aircraft target detection in zoom sensing images by the YOLOv4 algorithm. However, targets with large occlusions in bionic vision system cannot be detected accurately. In general, compared with other algorithms, YOLO series algorithms can meet the requirements of real-time and high efficiency in zoom sensing image detection. Compared with YOLOv4, YOLOv5 had higher accuracy and no reduction in detection speed.

Because the YOLO algorithm has good comprehensive performance in detection speed and accuracy, it is applied to short-range target detection. However, it is difficult to adapt to the huge difference of target size and scale in target detection (Liu, Y. et al., 2021; Yun et al., 2022). According to the structure of curved zoom microlens array, the single-stage sub eye lens of bionic human eye imaging system is a microlens array with uneven focal length. Its transverse image points are in the plane of photoelectric receiver, while the longitudinal image points will have obvious defocus. Therefore, the detection effect become better after adjusting the vertical grid. Based on YOLO deep learning network and the characteristics of image distortion caused by the special imaging principle of bionic zoom vision system, a new real-time detection method of zoom sensing targets is proposed in this paper.

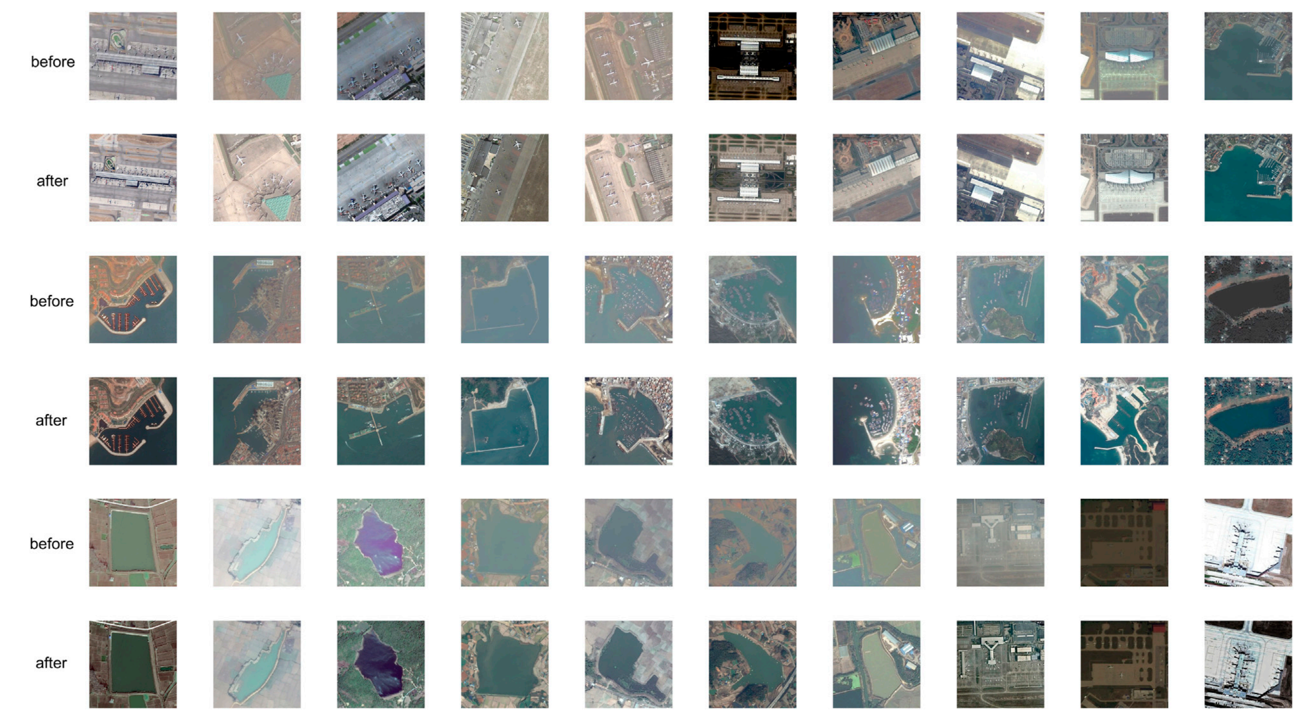

FIGURE 3 | Comparison of light and dark scenes before and after separation.

ZOOM TARGET DETECTION BIONIC VISION METHOD

Zoom Image Texture Feature Detection Index

The information of high-resolution images is diversified and complex. For such data sources, information processing technology cannot stay on several simple technologies. For the determination of the target area, in general, if the scope of the target is fixed or has access to the area of specific geographic coordinates, GPS and GIS technology can be used to guide the satellite to the specified area which is most intriguing, and fixed location monitoring of the data search can greatly improve the efficiency and precision of target detection (Wu et al., 2020; Yang et al., 2021; Zhao et al., 2022). The overall processing of zoom sensing images is shown in **Figure 1**. First, the high-resolution zoom sensing image is preprocessed, such as geometric correction and noise filtering. The generated image is extracted for information, usually spectral and texture features. The output results, such as ground feature classification maps, can be further used in practical application scenes, such as road extraction (Al-Masni et al., 2018; Li and He, 2020).

A Texture expresses the local properties of the image and reflects the relationship between pixels in the local area. It is generally believed that texture is a grey co-occurrence matrix composed of texture elements repeatedly arranged according to certain rules. It is defined as the probability of n grey at the x' point (i', j'), which is displaced d away from (i, j) in any direction, when the grey level at pixel (i, j) of point x in the image is M. The grey cooccurrence matrix can be used to extract feature parameters to describe the difference in ground object spatial features. Besides, the richness of texture information quantitatively describes the texture features of remote sensing images.

Different from image features such as gray and color, texture is represented by the gray distribution of pixels and their surrounding spatial neighborhood, which is represented as the local texture information. While various parameters of texture features reflect the properties of global features, they also describe the surface properties of the scene corresponding to the image or image region. The four features of images are entropy (ENT), HomoM (HOM), contrast (CON), and angular second moment (ASM). The synergy mainly reflects the degree of local variation of the image texture, that is, the uniformity of the image. Contrast can be used to reflect the clarity of the texture or the quality of the visual effect of the image. Information entropy can measure the richness of image texture information. The angle second moment is mainly used to observe the thickness of the image texture. The specific calculation formulas are as follows:

$$ENT = -\sum_{i,j} p(i, j)^2 \log_2 p(i, j) \tag{1}$$

$$CON = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} |i - j|^2 p(i, j) \tag{2}$$

$$HOM = \sum_{i,j} \frac{p(i, j)}{(i - j)^2 + 1} \tag{3}$$

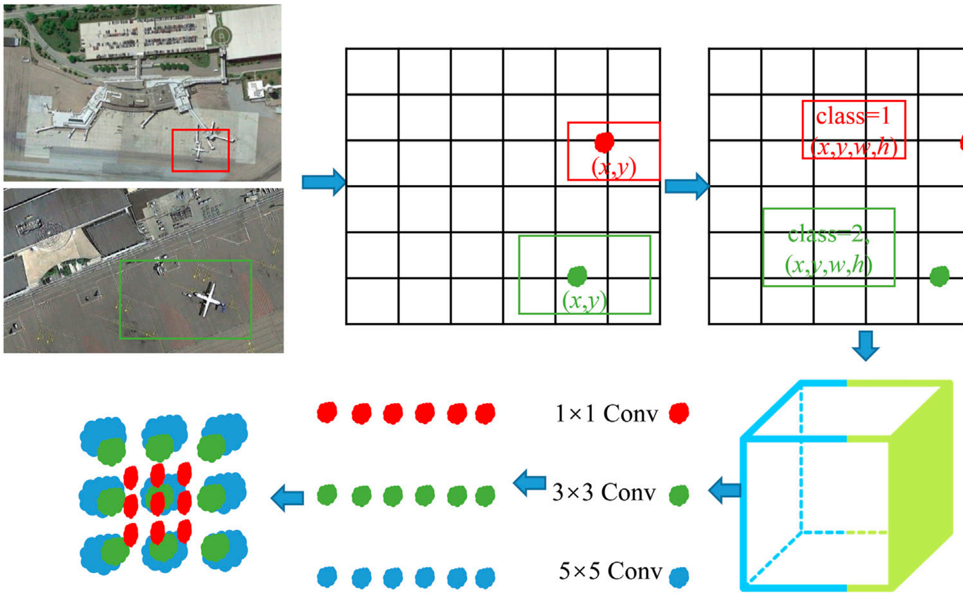

FIGURE 4 | Detection principle of the YOLO model.

$$ASM = \sum_i \sum_j p(i,j)^2 \tag{4}$$

Where $p\ (i, j)$ is the element in the normalized grey cooccurrence matrix; i and j are the row and column numbers of pixel points; and L is the grey level. The distortion of biomimetic remote sensing images will damage their texture features, so we use texture features to detect the distortion of biomimetic remote sensing images.

Bionic Sensing Image Correction Model

Due to the problem that remote sensing images cannot extract effective information caused by dark shadows, a TMF (truncated media filter) is applied which not only remove noise but also enhance the boundary. First, the pixels with the same characteristics in the dark part of the image are labelled, and the pixels of the labels are connected to form the dark part boundary. The image expression formed by the dark part pixels in the boundary region is as follows:

$$BndCon(T) = \left|\{q|q \in T, q \in Bnd\}\right| \sqrt{\left|\{q|q \in T\}\right|}^{-1} \tag{5}$$

In the formula, Bnd is the collection of dark boundary pixels, and q is the image fragment of the dark region. The TMF model detects the target by analyzing the associated parameters of region T and image light and dark pixels, which can realize the rapid identification and separation of image between light and dark scenes. The TMF model is shown in **Figure 2**.

The SLIC super pixel segmentation method is adopted in the model (Kumar and Bhandari et al., 2021; Chen et al., 2022) to transform the color image into 5-dimensional feature vectors in CIELAB color space and XY coordinates and then construct distance metric standards for the 5-dimensional feature vectors to perform local clustering of image pixels. The dark part of the image is converted into a super pixel image array, and the adjacent pixels in the super pixel array are released regularly so that there is a similarity relationship between them; thus, the shortest distance between the similarity degrees can be calculated as follows:

$$f_{hrp}(q,w) = \min_{q_1=q,q_2,\ldots,q_i=w} \sum_{n=1}^{i-1} f_{sqq}(q_n, q_{n+1}) \tag{6}$$

Where hrp stands for the shortest similarity distance from q to w, f_{hrp} represents the number of similarity systems between dark fragments q and w of super pixels, and f_{sqq} represents the Euclidean distance of colors in a super pixel abstract space. According to the calculation results, the block area of the dark part image can be obtained as follows:

$$Area(q) = \sum_{n=1}^{l} \exp\left[-\frac{f_{hrp}^2(q,q_n)}{2S_{vat}}\right] = \sum_{n=1}^{l} D(q,q_n) \tag{7}$$

The fusion value of the contour perimeter and dark pixel fragment area is calculated as follows:

$$BndCon(q) = \frac{Len_{bnd}(q)}{\sqrt{Area(q)}} \tag{8}$$

On normal conditions, the boundary shading values of the target pixel and the background pixel are different. The boundary shading values of the target pixel tend to be closer to 0, while the boundary shading values of the background pixel tend to be closer to 1. Therefore, a self-defined segmentation value can be used to achieve the segmentation of light and dark scenes and obtain the saliency region (Sun et al., 2022). When the impulse noise

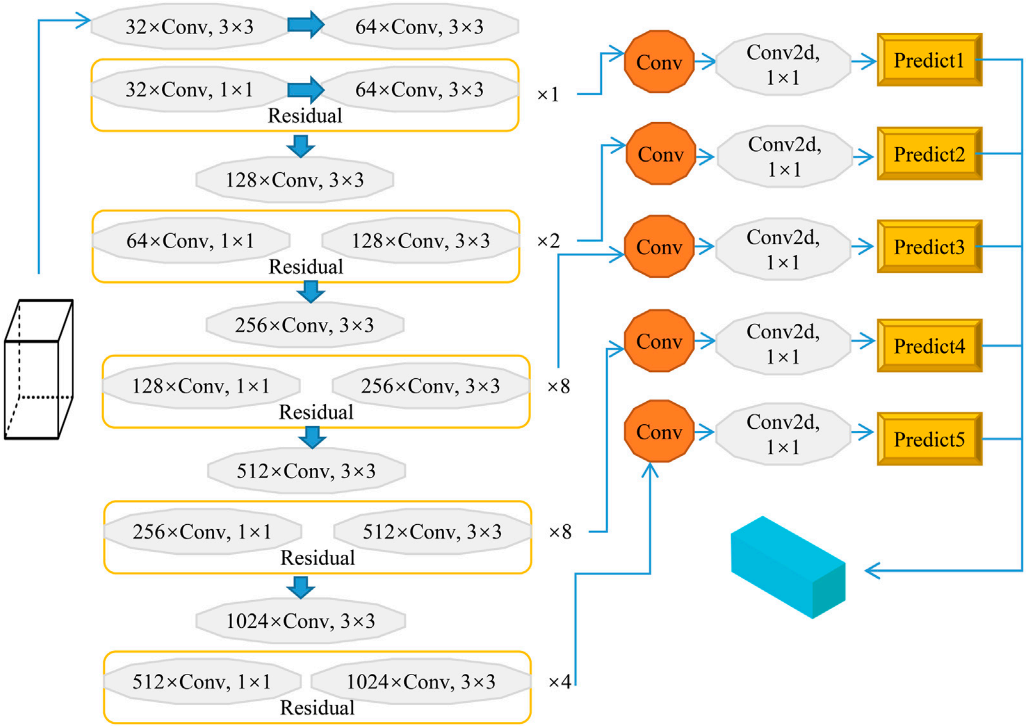

FIGURE 5 | Improved YOLO model for high spatial resolution biomimetic remote sensing target detection.

is very large, the TMF suppression effect is very outstanding. **Figure 3** is the effect diagram after the separation of light and dark scenes. The algorithm separates the target white scene from the background dark scene and effectively enhances the edge. After the above calculation, the image can basically reach the recognition standard.

Principles of the YOLO Algorithm

The first step of traditional target detection is to extract features, such as the LBP feature and HOG feature. Secondly, object model can be trained by SVM approach, then matches the model with the target. YOLO uses the idea of regression to integrate target region prediction and target category prediction into a neural network model. In this method, a neural network is used to unify candidate frame extraction, feature extraction, target classification and target location to achieve end-to-end target detection.

Biomimetic sensing images may contain multiple targets or categories of targets, so it is necessary to judge multiple categories of each prediction frame. The complete detection process of YOLO is shown in **Figure 4**. The specific detection process is as follows:

1) The image is divided into $S \times S$ grids. Each grid is given B prediction frames. Each prediction frame contains 5-dimensional information, namely, (x, y, w, h, c), where (x, y) is the predicting boundary centre relative to the offset of cell boundaries, (w, h) is the boundary of the width and height relative to the proportion of the whole image, and c is an incredible value, that is, the confidence that the target is included in the prediction boundary, as shown in **Formula (9)**. Here, $Pr(o)$ indicates whether the target exists in the cell corresponding to the prediction frame, one indicates existence, and 0 indicates nonexistence. IoU is the intersection ratio between the prediction box and the real value.

$$c = \Pr(o) \times IoU \tag{9}$$

2) CNN extract characteristics and prediction of each grid are presented when an object of class c has the conditional probability $Pr\ (c|o)$, and then the probability of each category in the network is obtained. The probability of a certain class is multiplied by the corresponding confidence, and the confidence value of the class is obtained, as shown in **Formula (10)**:

$$c = \Pr(c|o) \times \Pr(o) \times IoU = \Pr(c) \times IoU \tag{10}$$

3) The filter frame is suppressed according to the non-maximum value, then output the final judgement result. To optimize the model, YOLO uses the $S \times S \times (B \times 5 + c)$ dimension vector and mean square error of the image truth value as the parameters of the loss function. However, since there are no target objects in many meshes, different scale factors are set to balance the predicted boundaries regardless of the existence of targets when designing the loss function of YOLO, and the loss factors of boundary boxes need to be distinguished.

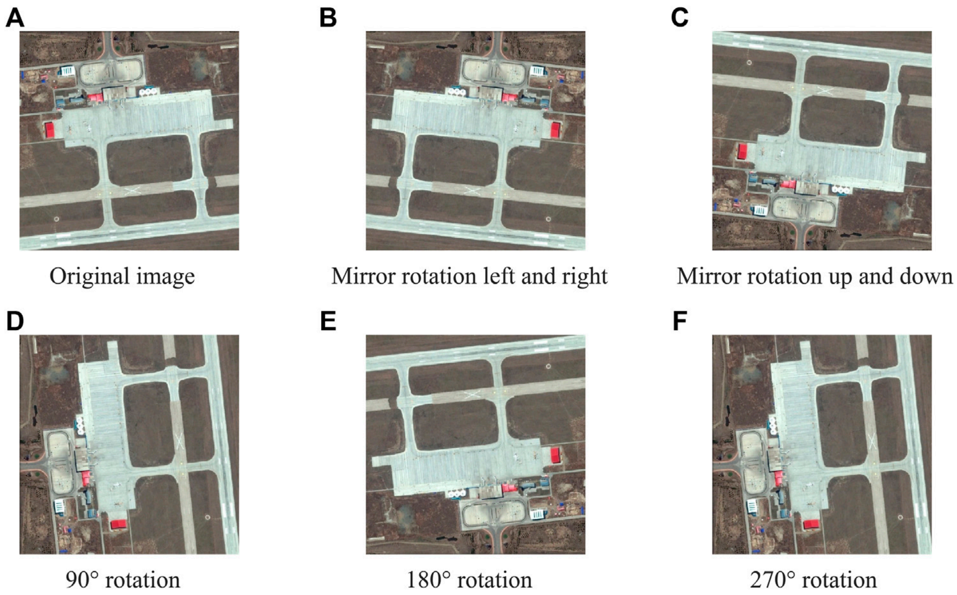

FIGURE 6 | Dataset enhancement example.

TABLE 1 | Selection of key parameters.

Batch size	Epoch number	Momentum	Decay	Learning rate
2	300	0.937	0.0005	0.01

For the loss coefficient of category judgement, for example, the loss weight of the boundary box is set to be 10 times the loss coefficient of category judgement. The above design gives the loss coefficient of the boundary box a higher weight. The form of the loss function is shown in **Formula (11)**.

$$l = \alpha \sum_{i=0}^{S^2} \sum_{j=0}^{B} \left[(x_i - \bar{x})^2 + (y_i - \bar{y})^2 + \left(x_j - \bar{x}\right)^2 + \left(y_j - \bar{y}\right)^2 \right] \tag{11}$$

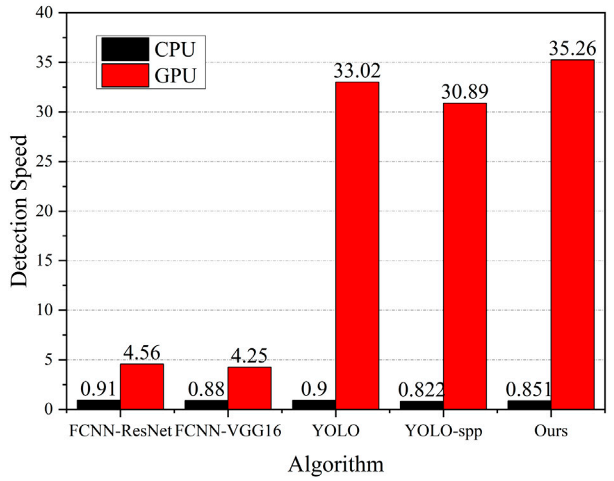

FIGURE 7 | Comparison of detection speed.

Addition of a Vertical Grid to Improve the YOLO Network

In the YOLO detection method, images are divided into $S \times S$ grids, that is, horizontal and vertical detection weights are the same. However, the length-to-width ratio of the target in the enlarged image obtained after biomimetic remote sensing image correction cannot accurately reflect its true value. The nonlinear target deformation and different deformation density in the same direction showing a phenomenon of imbalance between the upper and lower parts.

Taking the spherical center of the curved zoom of the bionic human eye as the origin of the coordinate system into consideration, if we want to use the linear model to locate the target, what need to obtain the positions of two or more pixels of the same target to determine the three-dimensional world coordinates of the target point in space. The camera passes through the lens imaging system, and its projection can convert the three-dimensional scene into the two-dimensional plane of the camera through imaging transformation. To solve this problem, the predicted frames in the vertical axis direction are altered. In this paper, the vertical number is doubled without changing the horizontal number; that is, the grid number is changed from S × S to S × 2S, and it is increased at the end of the

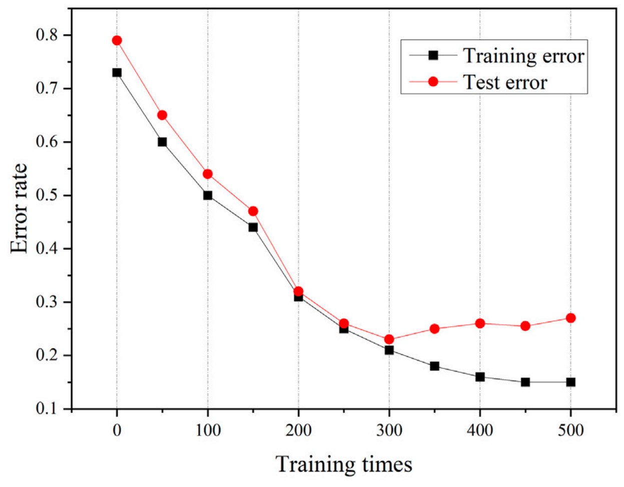

FIGURE 8 | Variation curves of training error and test error over time.

YOLO network structure. This network structure includes 20 convolutional layers and five maximum set layers and generates the improved YOLO network, as shown in **Figure 5**, to meet the needs of high spatial resolution remote sensing image detection.

EXPERIMENTS AND RESULTS

Dataset Pre-processing

The application scenario of the experiment in this paper is target detection from the perspective of high-resolution zoom sensing. Therefore, the dataset used must have the characteristics of zoom sensing images, and the common characteristics are as follows: 1) This is one of the most distinctive characteristics of biomimetic remote sensing images, which means the image is acquired from a high-altitude view, presenting a picture effect of the bird's eye view. Therefore, the detector developed by conventional data training cannot achieve ideal results. 2) Scale diversity. As the flight height is not fixed when acquiring data, the shooting height of remote sensing images is uneven, which results in different sizes of ground targets, furthermore even similar targets will show different sizes. 3) Target complexity. This is the difficulty of zoom target detection training. Since the most targets in this scene are small targets composed of dozens of or even several pixels, feature information is scarce. In addition, remote sensing images contain too many targets, as well as different target states and directions, resulting in very complicated and difficult training. 4) Background interference. Zoom sensing images usually cover a coverage area of several square kilometers, with a large field of vision. Therefore, the background often contains rich information, such as plain green land, mountains, rivers and road courses, with strong interference.

To make the data meet the above characteristics, the collection of datasets needs a special aerial zoom sensing image database. Commonly used open datasets include aerial image dataset (AIDDataset), DOTA, UCAS-AOD, NWPUBHR-10, etc. The AIDDataset used in this paper is a remote sensing image dataset that contains scene images of 30 categories, of which each category has approximately 220–420 images, and the overall total is 10000 images, of which each pixel size is approximately 600 × 600. The dataset, published by Huazhong University of Science and Technology and Wuhan University in 2017, contains

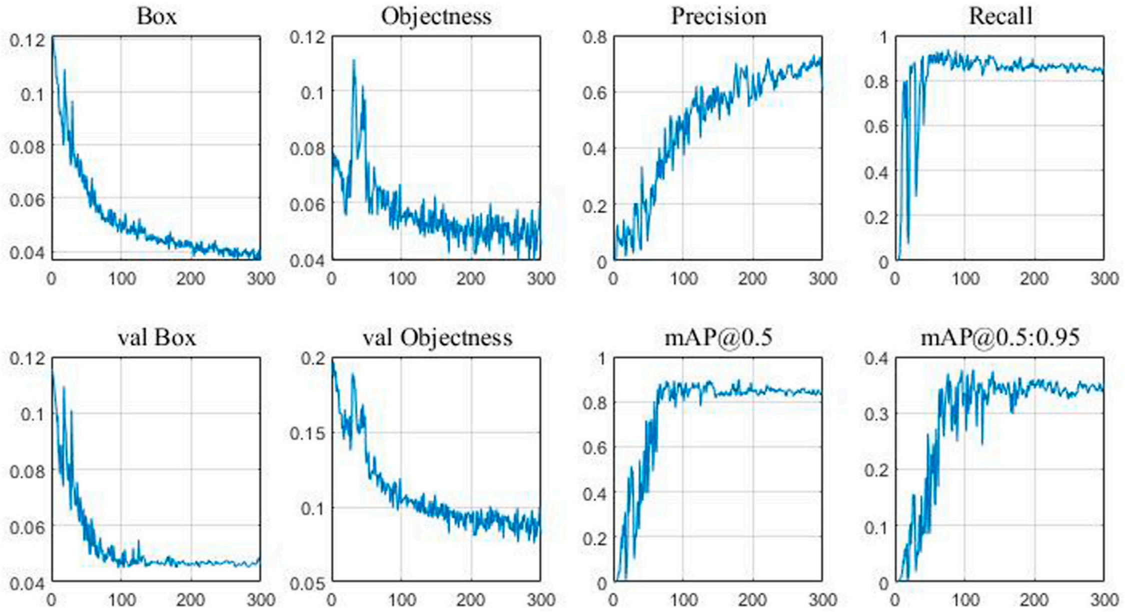

FIGURE 9 | Visualization of training results.

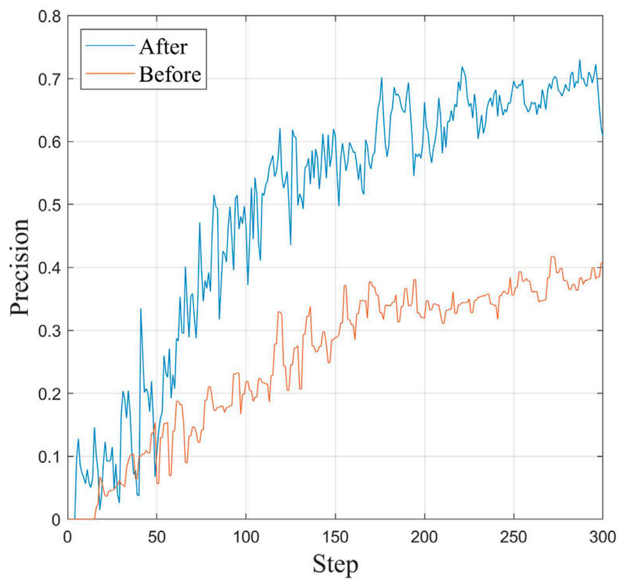

FIGURE 10 | The improved YOLO algorithm and the change in accuracy with training times before improvement.

TABLE 2 | Test results of different methods.

	mAP	Recall	Avg time
FCNN(ResNet101)	0.862	0.931	252.21
FCNN(VGG16)	0.874	0.922	164.16
YOLO	0.767	0.959	28.67
YOLO-spp	0.822	0.905	25.23
Our	0.851	0.972	29.5

TABLE 3 | Detection results of a single target by different methods.

Parking	Port	Storage tank	Airplane
0.468	0.731	0.893	0.901
0.423	0.732	0.858	0.908
0.439	0.657	0.765	0.825
0.503	0.692	0.906	0.916
0.876	0.852	0.901	0.919

a wide range of images under different imaging conditions, with different sizes of the ground and objects in the images.

Preliminary processing of the dataset is as follows: First, the TIF format data of zoom sensing images after multispectral and panchromatic data fusion are converted to PNG format, which is commonly used in network models. Then, the training data and test data suitable for network model training are screened. The dataset postprocessing includes five stages: data annotation, mask generation, sample cutting, down sampling and data enhancement. Because there is too much texture information in high-resolution zoom sensing images, it will affect the judgement of the network model and reduce the segmentation accuracy. Therefore, the sample size of 600 × 600 was reduced to 256 × 256. To enrich the image training set, better extract the training features, and generalize the model (to prevent overfitting), image block rotation, distortion, increase of noise and other ways to enhance the image were utilized. Specifically, the original image is flipped 90°, 180°, and 270°, and it is mirrored horizontally and vertically to generate another five pieces of data. The original data are added, and the amount of data increases by 6 times. The same operation is performed on the label, as shown in **Figure 6**. Through the above operation, the new training set contains 12000 pieces 600 × 600 sub images. To evaluate the training effect of the proposed network model on different datasets, the data set used for the test is divided into test set 1 (20000 pieces, 256 × 256), testing set 2 (20000 pieces, 512 × 512) and testing set 3 (20000 pieces, 600 × 600).

Algorithm Setting and Detection Speed

The training and testing of the network model are carried out in the server of the laboratory. The configuration of the computer is Windows 10 system, the CPU is Intel core5-7200u, the GPU is NVIDIA GeForce GTX940, and the memory is 8G. Cuda10.2, python3.6, and the corresponding extensions are installed.

Torch can be applied to a wide range of algorithms for machine learning. Compared to Caffe, the interface is easier to use, and the bottom layer is the C/CUDA execution program. PyTorch, a version of Torch, is a Python-first deep learning framework that uses a powerful GPU such as Numpy to speed up tensor calculations. PyTorch's interface is flexible and easy-using, with more excellent performance than other frameworks. In summary, considering the environment required for this experiment, the PyTorch framework is adopted in this paper as the platform for YOLO zoom sensing target detection by comparing the characteristics, performance requirements, deployment conditions and ease of use of the network model.

Limited by the laboratory computer configuration, the batch size was 2, in which the default optimal value of the YOLOv5 algorithm was adopted for momentum. Some parameters during training were selected, as shown in **Table 1**, and experiments were carried out after parameter setting was completed.

The improved algorithm and other algorithms were tested on a CPU and GPU, respectively. The statistical results are shown in **Figure 7**. The detection speed of the algorithm is calculated according to the average number of test images per second. The more pictures processed, the faster the detection speed of the algorithm.

No CPU can meet the real-time requirements of zoom sensing detection. The speed of using YOLO or the proposed algorithm on a GPU is more than 30 pieces per second, which far meets the detection speed requirements. The current mainstream detection method, Faster-RCNN, has a higher accuracy rate, but the detection speed is less than five frames per second. It embodies the excellent performance of YOLO and the improved YOLO algorithm in zoom sensing real-time detection.

The curve of training error over time is shown in **Figure 8**. It can be seen from the figure that the training error continuously decreases with the progress of training and finally reaches the lowest error rate of 21%. However, the test error reaches the lowest error rate of 23% when the training reaches 300 times, and then the error rate increases. Therefore, the optimal classification model finally obtained is the model trained 300 times. The model parameters obtained at the 300th

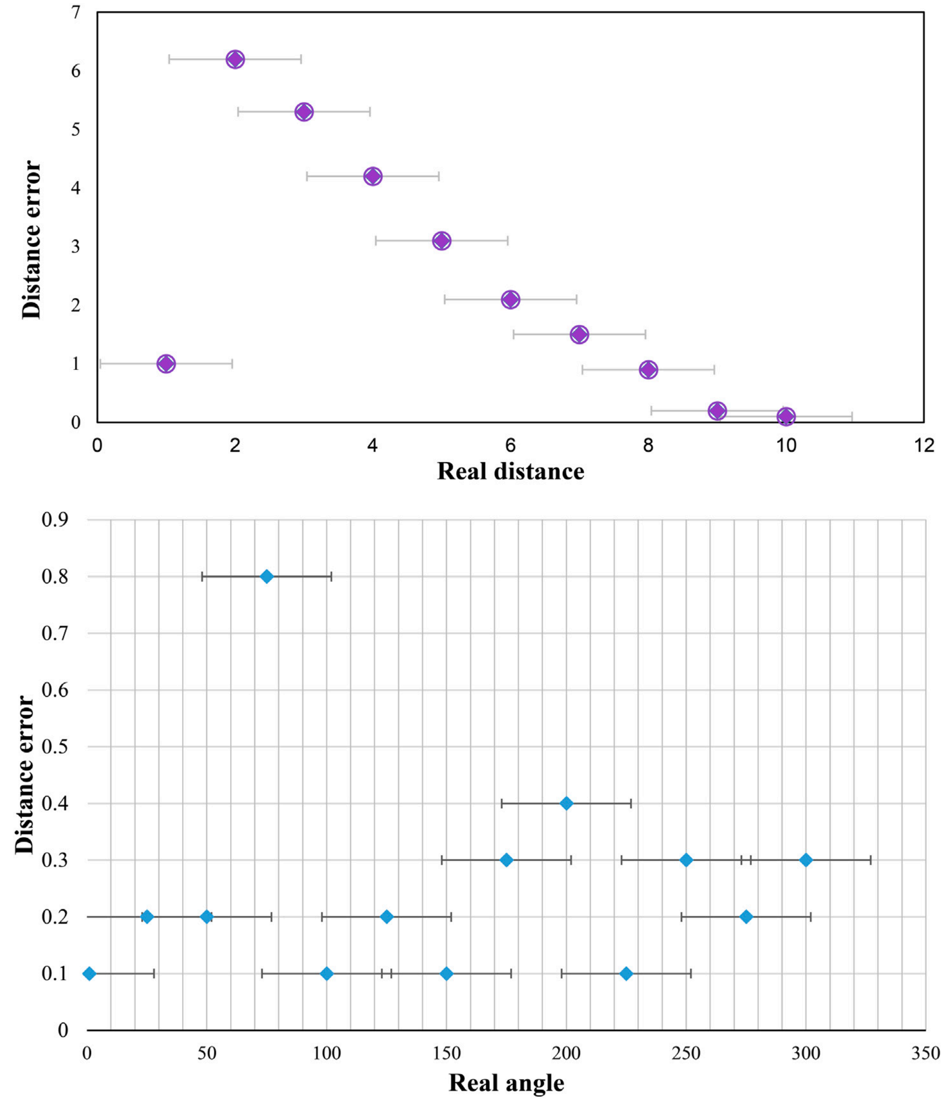

FIGURE 11 | Static distance and angle error.

training session are shown in **Figure 9**. YOLOv5 uses GIOU loss as the loss of bounding box. Box is the mean value of GIOU loss function. The smaller the value, the more accurate the box is. Objectness is speculated as the mean value of target detection loss. The smaller the target, the more accurate the target detection is. Classification is the mean value of classification loss. The smaller the classification, the more accurate the classification is. Precision indicates that the correct positive class is found. Recall means the true positive accuracy, that is, how many positive samples have been found. From the perspective of recall, the number of real classifiers is described, that is, how many real classifiers are recalled. Val BOX is the verification set bounding box. Val Objectness is the mean value of target detection loss in the verification set. Val Classification is the mean value of the classification loss of the verification set. mAP is the area surrounded by precision and recall as two axes, m represents the average, and the number after at represents the threshold for judging IoU as positive and negative samples.

Comparison of Accuracy

The accuracy of target detection can be indirectly reflected by the detection error rate. The lower the error rate is, the more reliable the detection accuracy of the corresponding model is, as shown in **Figure 10**. The following conclusions can be drawn.

1) The zoom sensing image is directly detected without preprocessing, the detection effect is very unsatisfactory, and it has lost practical value. In this paper, the YOLO expansion diagram is used for detection, and the error rate is approximately 35%. Because the detected target is in a "short and thick" state, traditional detectors cannot adapt to it. Using the algorithm in this paper, by increasing the number of longitudinal prediction frames to adapt to short and thick targets, the error rate is controlled at approximately 30%, and the performance is far better than that of YOLO.

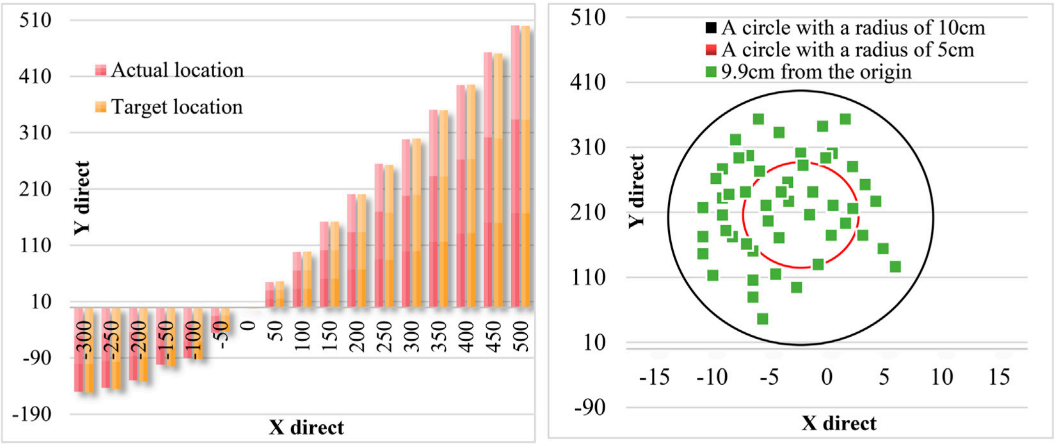

FIGURE 12 | Position comparison diagram and static error diagram.

2) Compared with Faster-RCNN and other detectors, the error rate of Faster-RCNN is less than 30%, which is the detector with the best accuracy. The error rate of the proposed algorithm is slightly higher, approximately 3%, but considering the huge advantage of the detection speed of the proposed algorithm, it is more practical in high spatial resolution zoom sensing image detection.

The phenomenon of misclassification and missing classification is obvious in the traditional Faster-CNN method, and the edge details are rough and messy, so it cannot effectively identify the real object category in the shaded area, and it cannot accurately classify small targets such as aircraft. This is because traditional classification methods are restricted by many factors, such as the image segmentation scale and classifier performance. The deep learning method FasterCNN-VGG16 has fewer errors and omissions, but it is not effective in distinguishing low vegetation and trees and processing details such as building edges and vehicles. This is due to the phenomenon of large differences within the class of high-resolution zoom sensing images, i.e., the same object with different spectra or foreign objects with the same spectrum. In addition, it can be seen from the classification results of the two groups of experiments that the classification effect of existing deep learning methods is not stable. The classification effect of the proposed method is the best. After the TMP model, the light and dark scenes are separated, the edges are smooth and accurate, and the details of ground objects are reflected more comprehensively and truly.

YOLO, YOLO-spp and the improved YOLO network were trained, and the detection performance of each model was tested on the test set. To fully demonstrate the performance of the improved algorithm, Faster-RCNN and other networks are added in the experiment for comparative analysis. The comparison of detection results is shown in **Table 2**.

As seen from the data in the table, the average accuracy of the improved YOLO model for all targets on the test set is 85.1%, and the recall rate is 97.2%, which is increased by 8.4 and 1.3%, respectively, compared with YOLO. The accuracy of all single targets was improved. The mAP values of YOLO and other detection algorithms in its series, such as YOLO-spp, are 76.7 and 82.2%, respectively, and the recall rates are 95.9 and 90.5%, both lower than that of the improved YOLO model. For the detection results of a single target, as shown in **Table 3**, the map value of any other algorithm, whether in parking, port, storage tank or airplane, is significantly less than the optimized Yolo algorithm used in this paper, which can prove that this algorithm has a good recognition rate in zoom sensors images.

Angle Error Test

Figure 11 shows the angle error of the zoom sensing image obtained by the drone. **Figure 11** shows that the mean error and standard deviation of targets at different angles are basically the same, which indicates that the improved YOLO model has good detection accuracy.

Figure 12 is the comparison diagram and static error diagram of the observed actual position and the position obtained by the visual system. As seen from the figure, the maximum static distance error of the visual system is 9.9 cm, which is not much different from the real-time detection error.

Experimental results show that the improved YOLO algorithm not only has good comprehensive performance in terms of detection speed and detection accuracy but can also meet the detection requirements when adapting to target differences at different angles in zoom sensing target detection.

CONCLUSION

In this paper, the resolution of bionic sensing images and target detection accuracy are improved by optimizing some parameters of texture features in the light and dark areas of zoom images. A target detection method based on the improved YOLO algorithm is proposed. Experimental results show that the average precision of this method for all targets images on the test set in the GPU environment is 85.10%. Finally, compared with traditional YOLO algorithm, the precision and recall rate are increased by 10.91 and 1.40%, respectively, and the detection speed is increased by 6.78%.

In addition, the average error and standard deviation of targets at different angles are basically the same with good detection accuracy, which meets the target detection requirements of high-resolution zoom sensing images, and the accuracy of target recognition is maintained at greater than 70%.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

AUTHOR CONTRIBUTIONS

XS was responsible for manuscript preparation and worked as a supervisor for all procedures. GS was responsible for programming and data processing. HR and WZ participated in discussions and revisions. All authors contributed to the article and approved the submitted version.

FUNDING

This work was Funded by the China Postdoctoral Science Foundation (No. 2021M692473), Natural Science Foundation of Anhui Province (2108085QF260), The Open Research Fund of National Engineering Research Center for Agro-Ecological Big Data Analysis & Application, Anhui University (No. AE202103), Anhui Provincial Department of Education Research Project (KJ 2021A0179), the Key Natural Science Research Project for Colleges and Universities of Anhui Province KJ2021ZD0056, and Independent Project of Anhui Key Laboratory of Smart Agricultural Technology and Equipment (APKLSATE 2019X001).

REFERENCES

Abdollahi, A., Pradhan, B., Shukla, N., Chakraborty, S., and Alamri, A. (2020). Deep Learning Approaches Applied to Remote Sensing Datasets for Road Extraction: A State-Of-The-Art Review. *Remote Sens.* 12 (12), 1444. doi:10.3390/rs12091444

Al-Masni, M. A., Al-Antari, M. A., Park, J.-M., Gi, G., Kim, T.-Y., Rivera, P., et al. (2018). Simultaneous Detection and Classification of Breast Masses in Digital Mammograms via a Deep Learning YOLO-Based CAD System. *Comput. Methods Programs Biomed.* 157, 85–94. doi:10.1016/j.cmpb.2018.01.017

Alganci, U., Soydas, M., and Sertel, E. (2020). Comparative Research on Deep Learning Approaches for Airplane Detection from Very High-Resolution Satellite Images. *Remote Sens.* 12 (3), 458. doi:10.3390/rs12030458

Chen, B., Chen, X., Chen, F., Zhou, B., Xiao, W., Fu, W., et al. (2022). Integrated Early Fault Diagnosis Method Based on Direct Fast Iterative Filtering Decomposition and Effective Weighted Sparseness Kurtosis to Rolling Bearings. *Mech. Syst. Signal Process.* 171, 108897. doi:10.1016/j.ymssp.2022.108897

Chen, T., Yin, X., Yang, J., Cong, G., and Li, G. (2021). Modeling Multi-Dimensional Public Opinion Process Based on Complex Network Dynamics Model in the Context of Derived Topics. *Axioms* 10, 270. doi:10.3390/axioms10040270

Chen, Z., Huang, M., Zhu, D., and Altan, O. (2021). Integrating Remote Sensing and a Markov-FLUS Model to Simulate Future Land Use Changes in Hokkaido, Japan. *Remote Sens.* 13 (13), 2621. doi:10.3390/rs13132621

Duan, W., Maskey, S., Chaffe, P. L. B., Luo, P., He, B., Wu, Y., et al. (2021). Recent Advancement in Remote Sensing Technology for Hydrology Analysis and Water Resources Management. *Remote Sens.* 13 (6), 1097. doi:10.3390/rs13061097

Giri, P., Ng, K., and Phillips, W. (2019). Wireless Sensor Network System for Landslide Monitoring and Warning. *IEEE Trans. Instrum. Meas.* 68 (4), 1210–1220. doi:10.1109/TIM.2018.2861999

Haut, J. M., Fernandez-Beltran, R., Paoletti, M. E., Plaza, J., Plaza, A., and Pla, F. (2018). A New Deep Generative Network for Unsupervised Remote Sensing Single-Image Super-resolution. *IEEE Trans. Geosci. Remote Sens.* 56 (56), 6792–6810. doi:10.1109/TGRS.2018.2843525

Hou, T., and Jiang, Y. (2021). Application of Improved YOLOv4 in Remote Sensing Aircraft Target Detection. *Comput. Eng. Appl.* 12 (57), 224–230. doi:10.3778/j.issn.1002-8331.2011-0248

Hu, F., Xia, G.-S., Hu, J., and Zhang, L. (2015). Transferring Deep Convolutional Neural Networks for the Scene Classification of High-Resolution Remote Sensing Imagery. *Remote Sens.* 7 (7), 14680–14707. doi:10.3390/rs71114680

Huang, L., Chen, C., Yun, J., Sun, Y., Tian, J., Hao, Z., et al. (2022). Multi-scale Feature Fusion Convolutional Neural Network for Indoor Small Target Detection. *Front. Neurorobotics* 16, 881021. doi:10.3389/fnbot.2022.881021

Huang, L., Fu, Q., He, M., Jiang, D., and Hao, Z. (2021). Detection Algorithm of Safety Helmet Wearing Based on Deep Learning. *Concurr. Comput. Pract. Exper* 33 (13), e6234. doi:10.1002/CPE.6234

Islam, M. S., Uddin, M. A., and Hossain, M. A. (2021). Assessing the Dynamics of Land Cover and Shoreline Changes of Nijhum Dwip (Island) of Bangladesh Using Remote Sensing and GIS Techniques. *Regional Stud. Mar. Sci.* 41, 101578. doi:10.1016/j.rsma.2020.101578

Jiang, D., Li, G., Tan, C., Huang, L., Sun, Y., and Kong, J. (2021). Semantic Segmentation for Multiscale Target Based on Object Recognition Using the Improved Faster-RCNN Model. *Future Gener. Comput. Syst.* 123, 94–104. doi:10.1016/j.future.2021.04.019

Jin, X., Wang, Y., Zhang, H., Zhong, H., Liu, L., Wu, Q. M. J., et al. (2020). DM-RIS: Deep Multimodel Rail Inspection System with Improved MRF-GMM and CNN. *IEEE Trans. Instrum. Meas.* 69 (4), 1051–1065. doi:10.1109/TIM.2019.2909940

Khan, M. A., Akram, T., Zhang, Y.-D., and Sharif, M. (2021). Attributes Based Skin Lesion Detection and Recognition: A Mask RCNN and Transfer Learning-Based Deep Learning Framework. *Pattern Recognit. Lett.* 143, 58–66. doi:10.1016/j.patrec.2020.12.015

Kumar, M., and Bhandari, A. K. (2022). No-Reference Metric Optimization-Based Perceptually Invisible Image Enhancement. *IEEE Trans. Instrum. Meas.* 71, 1–10. doi:10.1109/TIM.2021.3132086

Lee, H.-C., and Ke, K.-H. (2018). Monitoring of Large-Area IoT Sensors Using a LoRa Wireless Mesh Network System: Design and Evaluation. *IEEE Trans. Instrum. Meas.* 67 (9), 2177–2187. doi:10.1109/TIM.2018.2814082

Li, H., Li, C., An, J., and Ren, J. (2019). Remote Sensing Image Target Detection Based on Improved Convolutional Neural Network with Attention Mechanism. *Chin. J. Image Graph.* 8 (24), 1400–1408. doi:10.118334/jig.180649

Li, J., and Liu, Z. (2020). Self-measurements of Point-Spread Function for Remote Sensing Optical Imaging Instruments. *IEEE Trans. Instrum. Meas.* 69 (6), 3679–3686. doi:10.1109/TIM.2019.2938639

Li, J., Shen, Y., and Yang, C. (2021). An Adversarial Generative Network for Crop Classification from Remote Sensing Timeseries Images. *Remote Sens.* 13 (13), 65. doi:10.3390/rs13010065

Li, W., and He, R. (2020). Aircraft Target Detection in Remote Sensing Images Based on Depth Neural Network. *Comput. Eng.* 7 (46), 268–276. doi:10.1186/s13638-018-1022-8

Liu, L. X., Li, B. W., Wang, Y. P., and Yang, J. Y. (2019). Remote Sensing Image Segmentation Based on Improved Canny Edge Detection. *Comput. Eng. Appl.* 12 (55), 54–58.

Liu, X., Jiang, D., Tao, B., Jiang, G., Sun, Y., Kong, J., et al. (2021). Genetic Algorithm-Based Trajectory Optimization for Digital Twin Robots. *Front. Bioeng. Biotechnol.* 9, 793782. doi:10.3389/fbioe.2021.793782

Liu, Y., Jiang, D., Yun, J., Sun, Y., Li, C., Jiang, G., et al. (2021). Self-tuning Control of Manipulator Positioning Based on Fuzzy PID and PSO Algorithm. *Front. Bioeng. Biotechnol.* 9, 817723. doi:10.3389/fbioe.2021.817723

Mahmoodi, J., and Salajeghe, A. (2019). A Classification Method Based on Optical Flow for Violence Detection. *Expert Syst. Appl.* 127, 121–127. doi:10.1016/j.eswa.2019.02.032

Peng, J., Wang, D., Liao, X., Shao, Q., Sun, Z., Yue, H., et al. (2020). Wild Animal Survey Using UAS Imagery and Deep Learning: Modified Faster R-CNN for Kiang Detection in Tibetan Plateau. *ISPRS J. Photogrammetry Remote Sens.* 169, 364–376. doi:10.1016/j.isprsjprs.2020.08.026

Roslim, M. H. M., Juraimi, A. S., Che'Ya, N. N., Sulaiman, N., Manaf, M. N. H. A., Ramli, Z., et al. (2021). Using Remote Sensing and an Unmanned Aerial System for Weed Management in Agricultural Crops: A Review. *Agronomy* 11 (11), 1809. doi:10.3390/agronomy11091809

Sadykova, D., Pernebayeva, D., Bagheri, M., and James, A. (2020). IN-YOLO: Real-Time Detection of Outdoor High Voltage Insulators Using UAV Imaging. *IEEE Trans. Power Deliv.* 35 (3), 1599–1601. doi:10.1109/TPWRD.2019.2944741

Shadrin, D., Menshchikov, A., Somov, A., Bornemann, G., Hauslage, J., and Fedorov, M. (2020). Enabling Precision Agriculture through Embedded Sensing with Artificial Intelligence. *IEEE Trans. Instrum. Meas.* 69 (7), 4103–4113. doi:10.1109/TIM.2019.2947125

Sun, W., Du, H., Ma, G., Shi, S., Zhang, X., and Wu, Y. (2020). Moving Vehicle Video Detection Combining ViBe and Inter-frame Difference. *Ijes* 12 (12), 371–379. doi:10.1504/IJES.2020.107042

Sun, Y., Zhao, Z., Jiang, D., Tong, X., Tao, B., Jiang, G., et al. (2022). Low-illumination Image Enhancement Algorithm Based on Improved Multi-Scale Retinex and ABC Algorithm Optimization. *Front. Bioeng. Biotechnol.* 10, 396. doi:10.3389/fbioe.2022.865820

Wang, Y., and Dou, W. (2019). A Fast Candidate Viewpoints Filtering Algorithm for Multiple Viewshed Site Planning. *Int. J. Geogr. Inf. Sci.* 34 (3), 448–463. doi:10.1080/13658816.2019.1664743

Wu, X., Jiang, D., Yun, J., Liu, X., Sun, Y., Tao, B., et al. (2020). Attitude Stabilization Control of Autonomous Underwater Vehicle Based on Decoupling Algorithm and PSO-ADRC. *Front. Bioeng. Biotechnol.* 10, 843020. doi:10.3389/fbioe.2022.843020

Yan, S., Jing, L., and Wang, H. (2021). A New Individual Tree Species Recognition Method Based on a Convolutional Neural Network and High-Spatial Resolution Remote Sensing Imagery. *Remote Sens.* 13 (13), 479. doi:10.3390/rs13030479

Yang, Z., Jiang, D., Sun, Y., Tao, B., Tong, X., Jiang, G., et al. (2021). Dynamic Gesture Recognition Using Surface EMG Signals Based on Multi-Stream Residual Network. *Front. Bioeng. Biotechnol.* 9, 779353. doi:10.3389/fbioe.2021.779353

Yu, L., Gao, W., Gao, W., R. Shamshiri, R., Tao, S., Ren, Y., et al. (2021). Review of Research Progress on Soil Moisture Sensor Technology. *Int. J. Agr. Biol. Eng.* 14 (4), 32–42. doi:10.25165/j.ijabe.20211404.6404

Yun, J., Sun, Y., Li, C., Jiang, D., Tao, B., Li, G., et al. (2022). Self-adjusting Force/bit Blending Control Based on Quantitative Factor-Scale Factor Fuzzy-PID Bit Control. *Alexandria Eng. J.* 61 (6), 4389–4397. doi:10.1016/j.aej.2021.09.067

Zhang, K., Yang, K., Li, S., and Chen, H.-B. (2019). A Difference-Based Local Contrast Method for Infrared Small Target Detection under Complex Background. *Ieee. Access.* 7, 105503–105513. doi:10.1109/ACCESS.2019.2932729

Zhao, G., Jiang, D., Liu, X., Tong, X., Sun, Y., Tao, B., et al. (2022). A Tandem Robotic Arm Inverse Kinematic Solution Based on an Improved Particle Swarm Algorithm. *Front. Bioeng. Biotech.* 10, 832829. doi:10.3389/fbioe.2022.832829

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.