Focus-Based Multi-Scale Approaches for Tiny Target Identification in Aerial Images

[1*]Dr.K.Chandramohan, [2]Mrs.P.Sathyasutha, [3]Ms.S.Dharshini, [4]Mr.K.Vijayprabakaran, [5]Dr.R.Umamaheswari, [6]Mr.K.Venkatesan

[1]Prof/CSE Gnanamani College of Technology
chandramohancse@gmail.com
[2]AP/CSE Gnanamani College of Technology
sudha.ssrs@gmail.com
[3]PG Student Gnanamani College of Technology
dharshasanthi235@gmail.com
[4]AP/CSE Gnanamani College of Technology
vijayviruz10@gmail.com
[5]Prof/CSE Gnanamani College of Technology
Umait1978@gmail.com
[6]AP/CSE Gnanamani College of Technology
kvenkatesanme@gmail.com

ABSTRACT

Small targets in remote sensing images have too few discriminative features, making them easily confused with background information and difficult to find, which reduces the detection accuracy when detecting aerial images using general target detection networks. To solve the above problems, they proposed a remote sensing small object detection network based on an attention mechanism and multi-scale feature fusion and named it. First, the detection head enhancement module is designed to enhance the representation of small object features by combining multi-scale feature fusion and an attention mechanism. Secondly, the Attention Mechanism Channel Cascade (AMCC) is designed to reduce redundant information in the feature layer and protect small objects from losing information during feature fusion. Then, the regularized Wasserstein distance is introduced and combined with the generalized cross-sum as the position regression loss function to improve the model optimization weights and regression box accuracy for small objects. Finally, a target detection layer is added to improve the ability to extract target features at different scales. Experimental results on the drone dataset VisDrone2021 and a self-made dataset show that AMCC is 2.4% and 3.2% higher than YOLOv5s, respectively, effectively improving the recognition accuracy of small objects.

Keywords: *Multi-Scale Approaches**,** Attention Mechanism Channel Cascade (AMCC), Normalized Wasserstein Distance (NWD).*

1. INTRODUCTION

A common approach for feature fusion is to merge features by concatenating channels of feature maps or by element-wise addition. Adding element-wise allows the feature map to contain more information while keeping the dimensionality the same, but is less computationally intensive than the cascaded approach. With the continuous advancement of technology, images captured by drones are now widely used in remote sensing imaging, agriculture, wildlife protection, disaster monitoring, etc. Although existing object detectors have made great progress in object detection in natural scenes, the following problems still remain when directly applying such general object detectors to remote sensing images: (i) If the drone's flight altitude is not constant, the scale of the same type of object will be different in the captured images. For example, this is the case with the images in the dataset. (ii) Small objects have problems such as a small number of effective pixels, limited feature representation, and being easily affected by the background. (iii) Loss functions based on Intersection over Union (IoU) variants are more sensitive to small object offsets than to large object offsets. This makes it difficult to spot small objects.

Figure 1: Aerial Image introduction diagram

Figure 1 shows that the large scale difference between objects in remote sensing images poses a great challenge for object detectors. Therefore, it is important to obtain a detection network that can detect objects at different scales. A common method to solve the object scale change is to establish multi-layer feature fusion, such as feature pyramid network (FPN) and feature fusion module path aggregation network (PANet), bidirectional feature pyramid network (Bi-FPN), adaptive spatial feature fusion (ASFF), and neural architecture search feature pyramid network (NAS-FPN). All of these are improvements based on FPN. However, small objects have a small number of effective pixels, and the feature information is lost after passing through the backbone network, so the model cannot correctly learn the important spatial and semantic feature information of small objects. Therefore, in order to reduce the loss of small object information, it is necessary to increase shallow branches and improve the resolution of the detection head's feature map.

In object detection, a regression loss function represents the degree of consistency between the model output box sizes and locations and the actual box sizes and locations. The regression loss function undergoes a smooth L1 loss to loss function based on the L1/L2 loss and the IoU variant that is currently commonly used. YOLOv5 uses GIoU as the location regression loss. GIoU is an improved version of IoU. Unlike IoU, which focuses only on overlapping areas, GIoU focuses not only on overlapping areas but also on non-overlapping areas, so it can more accurately reflect the degree of overlap between the two. However, this loss function is very sensitive to small object misalignments. It is not suitable for small objects because even a small misalignment of a small object can cause a large increase or decrease in the IoU value. Other scholars have used variations of IoU to solve the small object regression problem to some extent, but such loss functions are still not suitable for small objects. NWD based on 2D Gaussian distribution is designed to effectively alleviate the problem of low detection accuracy of small objects in general object detection networks, but it does not consider the advantages of IoU-based loss functions in detecting medium and large objects.

Since the detection head of YOLOv5 contains the final object classification information and the regression information of the object frame, the detection head has a significant impact on the detection of small objects. During the model training process, due to the weak feature representation ability and the small number of pixels of small objects, the detection head detects too few small objects, which hinders the optimization weighting of small objects. Therefore, it is very important to significantly improve the ability to represent foreground features.

By utilizing multi-scale feature fusion, can obtain features of different sizes, enlarge the recognition area, and enhance the description of small object features, thereby improving the small object detection performance of the model. Feature fusion is the combination of information from different scales or branches and is an important component of the target detection network structure. However, this method may not be the optimal approach for problems involving semantic mismatch between input feature maps or mismatch in perceptual fields. Solving the problem of object scale change and imbalance of feature information of small objects in remote sensing image detection model.

2. Related Work

Currently, sensor information is a hot research topic, and sensor information acquisition and mobile computing are already relatively mature. In the case of image sensors, due to the limitations of certain devices, image information becomes more difficult to process, and the images captured by the sensors have characteristics such as high noise, small objects, and blurred objects. Object detection has great practical value and application potential, and is the foundation of many vision algorithm tasks such as image recognition and object tracking. One is a two-stage target detection network based on Region-CNN (RCNN), Fast Regional Convolutional Network (Fast-RCNN), Faster-RCNN, and Region-based Fully Convolutional Network (R-FCN). Then, the target is classified and regressed. This method compresses global pixels through convolution and weights the original feature shadow pixels, reducing noise interference and improving the network's ability to grasp global key pixel information. Experiments, remote sensing ship image recognition experiments conducted on remote sensing images, datasets showed that the network structure improves the detection performance of small targets [1-3].

This type of network has high detection accuracy but low real-time performance. One type is the single-stage object detection network, represented by single-stage multi-box detector (SSD), YOLO series, fully convolutional single-stage object detector (FCOS), and Retina Net, which directly extracts the semantic and spatial features of objects and completes object classification and regression. This type of network has low overall performance but high immediacy and is widely used in various scenarios. Academically, there are two ways to define small objects: relative size and absolute size. In the relative size method, an object is considered small if its aspect ratio is 0.1 of the original image size. In the absolute size method, an object is considered small if its dimensions are less than 32 x 32. This article uses the definition of absolute size. Remote sensing images have complex backgrounds, few effective pixels for objects, and various sizes and shapes. Existing general object detectors have difficulty extracting accurate and effective feature information for classifying and identifying small objects. To address the challenge of inaccurate detection of small targets in the field of remote sensing, this paper leverages the research achievements of other researchers in two aspects: multi-scale fusion and attention mechanism. The deep feature layer contains rich semantic object information and a wide recognition area, while the shallow feature layer contains more fine-grained information, so the deep feature information and shallow feature information can be appropriately utilized to improve the accuracy of the small object detection model through multi-scale fusion. A small object detection model using a single-shot multi-frame detector (DFSSD) featuring dilated convolution and feature fusion is proposed. By expanding the receptive area of features, obtaining contextual information of features at different scales, and enhancing the semantic information of shallow features, the detection effect of remote sensing small objects can be improved to a certain extent [4-6].

designed an extended feature pyramid network (EFPN) specialized for small object detection, which includes a feature texture transfer (FTT) module, which extracts semantic information and texture features from the feature map of the FPN network and acts on the super-resolution feature map, effectively improving the representation ability of small object feature information while achieving high computational and storage efficiency. propose a multi-scale dynamically weighted feature fusion network, which adaptively assigns different weights to feature layers at different scales through network training, enhancing the contribution of shallow feature information in the whole network, thereby deriving a model for small object detection tasks. To solve the problems of false positives, poor anchor box regression performance and inability to detect small targets in traditional multi-scale target detection methods based on YOLOv4, a new target detection framework, Enhanced YOLOv4, is proposed. First, the improved BiFPN replaces the original PANet as the feature fusion module and achieves multi-scale feature fusion by sharing weights. Secondly, a channel attention mechanism (CAM) is incorporated in front of the detection head to highlight correlations between channels and draw more attention to small objects. Finally, to improve the anchor box regression effect and speed up the training of YOLOv4, improved the network training loss function and replaced the original CIoU with CDIoU. Experimental results using the DOTA dataset verify the effectiveness of our improvements. Our method achieved an mAP of 90.88% and a frame rate of 58.76 FPS without any significant change in detection speed [7-9].

With the rapid development of remote sensing technology, remote sensing target detection technology has been widely used in many fields such as intelligent detection, rescue, civil and military applications. Remote sensing target detection technology has attracted more and more attention from experts and scholars. In large-scale satellite images, objects such as roads and bridges have been successfully identified as targets. Nevertheless, the detection of small objects in remote sensing remains a never-ending and challenging task. Changes in the orientation of objects in remote sensing images can seriously affect the accuracy and efficiency of target detection methods. In addition, the background of remote sensing images is complex, and there may be various interferences such as high inter-class similarity between objects, position changes, occlusion, geometric distortion, etc., which require higher accuracy and speed of target detection methods. Traditional machine learning and deep learning techniques are the two main categories of target detection technology. Compared with earlier target detection methods such as HOG-Cascade, the target detection method based on convolutional neural network (CNN) has significant advantages in speed and accuracy. To implement feed-forward convolution calculations, CNNs use relatively thick convolutional layers. With the rapid development of CNN, object detection methods based on deep learning have become a research hotspot[10-12].

This manuscript is based on the original images of the original images, the small size of the distribution, and the density of the distribution. It effectively avoids questions that would result in missing key characteristic information about the target. First, we design an enhanced multi-scale feature fusion pyramid network, DSI-FPN. We optimize the depth-wise Sparable Convolution operator, Involution operator, space operator, and F PN+PAN network with fewer parameters and computational complexity to generate feature maps that are more useful in network detection tasks. The second is the proposal for a proper space-based attention authority, SCBAM. By introducing a unique and particular organization into the CBAM module, we can add non-station information to the interaction while retaining only the unique station information, thus breaking the limitations of the convolution kernel, expanding the wild of the receiver of the model, and UP the feature performance capability of the model. Thirdly, there may be a problem of insufficient computing power when deploying target inspections. Therefore, we propose a jointly supervised network knowledge steaming framework with a special critical layer. The accuracy of the SET calculation is very high, as it adjusts for teacher loss, teacher loss, teacher real loss, and teacher loss to adjust for raw online learning trends, and the accuracy of the students is also high. The effect of "network parameters and model size" is mitigated by "do". Finally, regarding the method of comparing his portraits, experimental results show that this method has excellent detection effect on small targets in remote sensing images under various lighting conditions, with a test accuracy of 43.9% and an original sensor accuracy of 7.4%. After knowledge evaporation, a reduction of 1/3 of the original data was made and accuracy dropped to 40.2% [13-15].

3. Materials and methods.

Our goal is to develop a simple method to predict channel parameters from 2D satellite or aerial images without the need for 3D models. Existing object detection networks can be broadly divided into two categories. The importance of finding the necessary boundary corners at the edge of the image to protect the corners of the image is pointed out. If the first image is found in the frame, the edge points of the image are adjusted to fit using boundary corner detection techniques, and the still image is saved to process the required information features. Multiple separation points were detected. The Attention Mechanism Channel Cascade (AMCC) algorithm leverages points, small changes and scales, rotations, and various modeling and quantization to achieve reasonably consistent indexing and information retrieval. This technique adapts the frame space to find the right image to use. In this way, the image boundary angle detection technique can clearly identify the input image presented within the image. Captured image wear occurs when an image is captured through a detected image point on a node that stores information statically. Given an image, if there is inter-image information that matches that image, the server will provide it to the client. If no match is found, the server image will not be saved. In both detection processes, feature points are found and located in an image based on correlation.

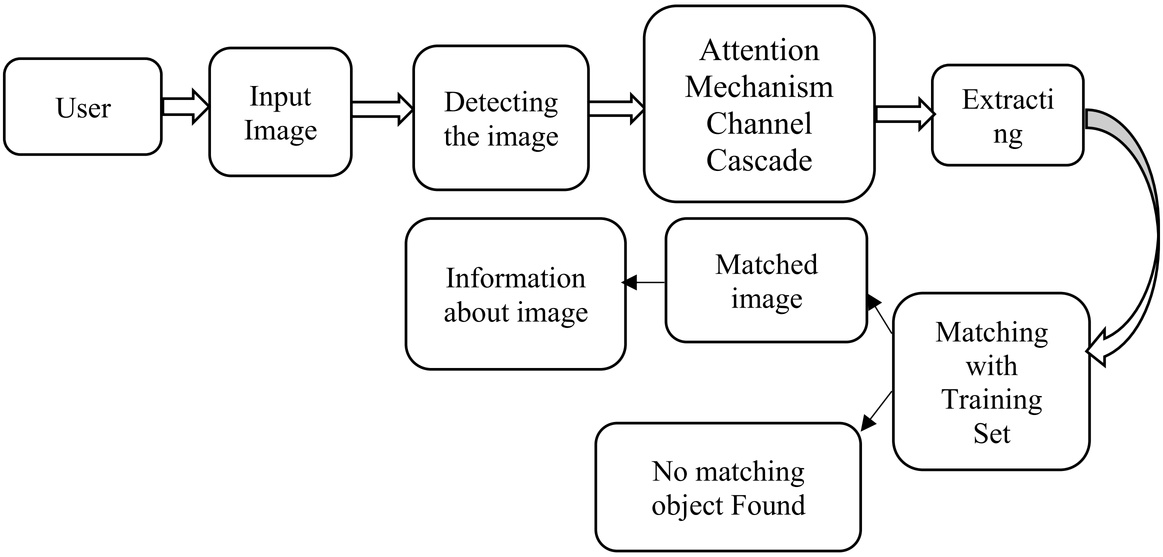

Figure 2: Implementation of the image detection diagram

This improves image detection based on corner points. If the partial constant strength is too small for both equations. If there is an edge, 1 is the smallest eigenvalue (since the gradient of the image is parallel, the eigenvalue is associated with the largest eigenvalue). Figure 2 shows the basis of border point image detection. The framework processes the client's input image. The Attention Mechanism Channel Cascade (AMCC) method is used to find and paint edge points in the image in the framework. The main drawback of most traditional machine learning methods is that they require a 3D model for prediction.

3.1 preprocessing of Image

There is a very elegant way to filter pre-processed samples. In this way, noise can be reduced through filtering. Video filtering is used to remove unwanted noise. To display frames updated in a way that eliminates changes in pixel values of objects, you can use only moving objects. The image rotates parallel to the horizontal direction of the line between the first two eyes.

Algorithm

Input: Text image dataset

Output: Preprocessed dataset

Step 1: Initialize images

Step 2: Load and resize images

Step 3: Process smoothed data

Journal of Harbin Engineering University
ISSN: 1006-7043
Vol 47 No. 06
June 2026

Step 4: For (Calculate luminance Ic for each image using heuristic saturation)
Calculate color histogram values of image objects
Apply and minimize peak signal-to-noise ratio (PSnr)
Identify each object representation and saturated image
End
Step 2: While (Saturated image (Gsi) □effort)
Smooth images using FCM □ Fast guide features
Step 3: Calculate gradient value map using GE
Step 4: Smooth using GE and normalize le images to range [0,255]
Step 5: Remove noise

Image features are resized to 256x256 pixels before cropping and feature extraction. Pre-processing is mainly to reduce the effect of noise on lighting, color intensity, background and orientation differences. Correct image recognition depends on the lighting conditions and the quality of the captured image. The recognition rate can be improved by preprocessing the captured images.

3.2 Persuasive points of Feature Extraction and its appearance

Due to the increased complexity of the system, including the same feature extraction methods and memory requirements for storing matched and non-matched features from the previous gender and age grouping stage, this stage uses corner estimation for selective learning since the appearance features extracted in the previous stage are still used. Regarding the quality of the relevant components, some principles are applied to evacuate the relevant components, such as most of them have long shapes and little noise components or stains appear on the report boundary due to their size, shape and position. Reduce noise by applying:

$$\text{Persuasive Boundary } F(x;m;s) = \frac{1}{x*s\sqrt{2\pi}} \exp - \frac{\text{In}(x-s)\text{^}2}{2s\text{^}2}$$

where x is a variable. m and s are the mean and standard deviation of the natural logarithms of all individual relevant component variables for a single image.

Feature extraction, feature simplification and feature classification for aerial images. Feature extraction for aerial images consists overanalyzing the biological relevance of the photos and, calculating the features.

3.3 Attention Mechanism Channel Cascade

This technique of finding the image that corresponds to the frame is used for matching. When the client captures the video, the Attention Mechanism Channel Cascade technique detects points in the given input image, changes the points to the detected edges of the image, and sends them to the server. This method proposes a theory for detecting image boundaries.

Multiple edge detectors are applied to the images, Palladian and gradient edge detection filters are applied to the images for visual comparison, and a soft transition edge detector is applied. Well, having a good performing edge detector doesn't necessarily result in a good image.

Step 1: Initialize edge detection and return Harris corner points.
Step 2: Detect similar points in image ← Erase unoccluded image (TC, edge image)
Step 3: Detect image ← (corner reduction image)
Step 4: Image gap change points ← Fill gaps (image cover image)
Step 5: TC ← Optimize coordinates (edge image, gap image, TC)
Step 6: Intent learning and transform comparison values ← Capture image (discrimination image, TC)
Step 7: Image ← Enhance contrast (match case image)
Step 8: Go back (match image)

In addition to corner algorithms, other edge detectors can also be applied. It is based on detecting the maximum transitions in different directions in the image. This edge detector uses a convolution matrix with image size 3x3. There are many variations of this convolution matrix, the main difference being in the rotation of the matrix. Four convolution matrices are applied to the image, and the boundary rotates the initial convolution matrix according to the maximum of four different directions.

3.4 Best match image

The extracted image matches the image stored on the server. If the image is matched, information about the specific image is displayed to the client. If the images don't match, the replay won't find a match. Many excellent results have been achieved in the research of basic data types of sensor information, and many relatively successful algorithms have been proposed.

4. Result and discussion

The results and performance of the proposed implementation are tested using the training function on the image dataset of the Mathematical Lab image processing tool. A performance evaluation was conducted to test the precision and reproducibility of the sensitivity and specificity measures obtained during the implementation phase. The test case measurements are calculated by comparing the true and false positions of the error rates performed during text processing. Performance is evaluated by measuring precision and recall on a set of positive and negative values in the test-training set.

Table 1 Details of Parameters Processing

Parameters used	Values processed
Input dataset	Arial Image dataset
Simulation tool	Python JupyterNotebook
Number of images	Training and trained images process

Table 1 details the natural image acquisition dataset that was processed to test the performance of the proposed system. Remote sensing images are often of low quality due to equipment limitations, resulting in low image accuracy. If an object is blurry or small, it becomes very difficult to identify it.

Table 2: Analysis of precision performance

Comparison methods/ No images	10 images	20images	30images
DWT	67.3	69.3	68.2
SVM	71.1	66.7	71.3
AMCC	74.3	74.1	75.2

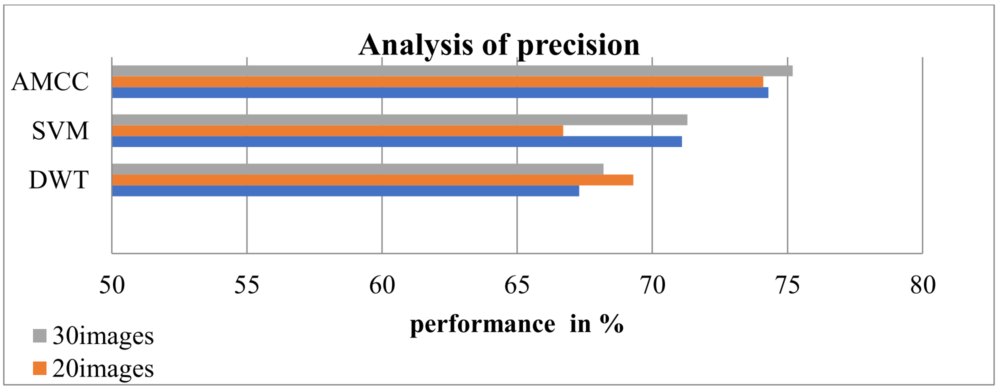

Figure 3: Analysis of a precision rate

Figure 3 and Table 2 show the true accuracy rates observed from different methods and different datasets, and the proposed implementation achieves higher efficiency rates than other methods. The results of cifar10 and cifar100 demonstrate the universality and practicality of the attention mechanism. The main challenge is to detect objects in images that have very few pixels.

Table 3: Analysis of recall rate performance

Comparison methods/ No images	10 images	20images	30images

DWT	67.3	66.3	68.7
SVM	75.2	71.4	73.2
AMCC	78.2	79.1	76.2

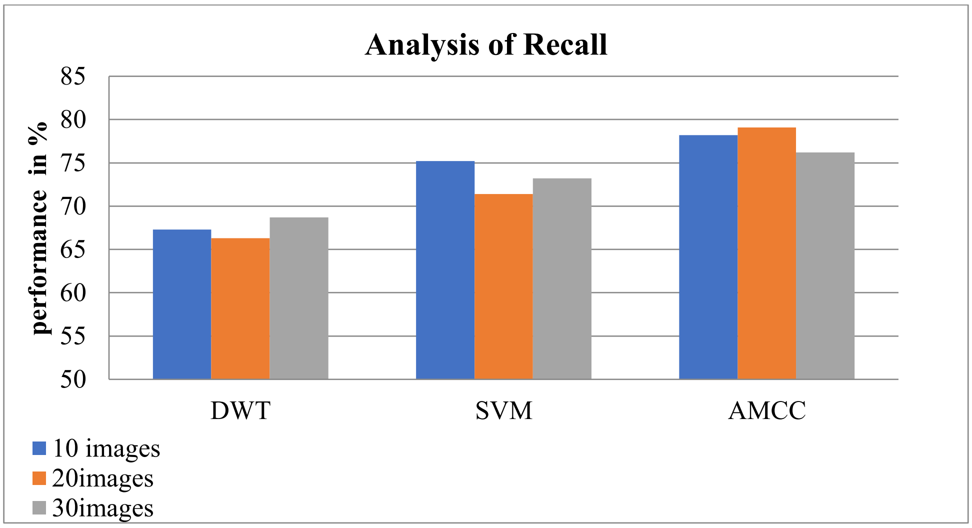

Figure 4: Analysis of recall rate

The recall analysis in Figure 4 and Table 3 above was tested using different images. The collected dataset has a variety of test values manufactured using different methods. The proposed system shows higher recall than other methods. Traditional convolutional networks extract information through local convolutions, which makes them more complex and susceptible to noise points, and therefore generally not suitable for small target classification and diagnosis.

Table 4: Analysis of detection accuracy performance

Comparison methods/ No images	10 images	20images	30images
DWT	91.2	94.2	95.3
SVM	92.1	94.6	95.6
AMCC	92.8	94.8	95.9

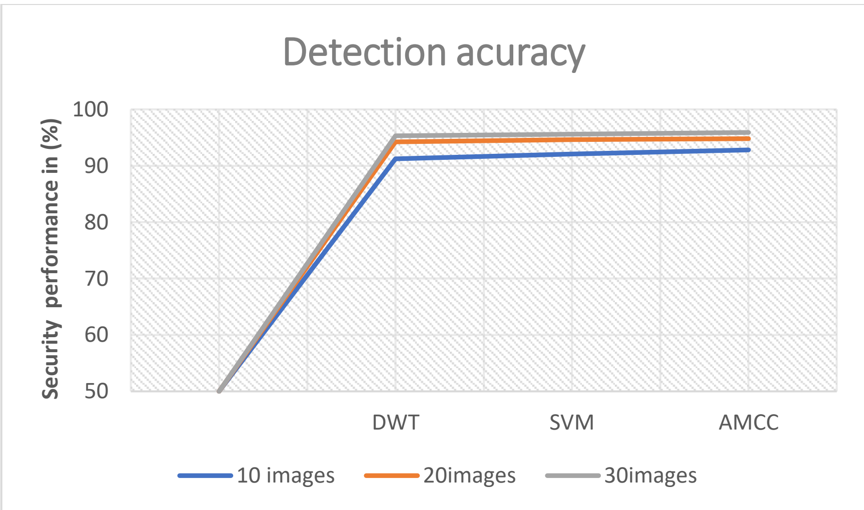

Figure 5: Comparison of security analysis

Another metric, shown in Figure 5 and Table 4, measures the level of accuracy with which users perform image detection. Compared with other methods, the proposed system has a significant impact on image detection performance. Current solutions involve processing feature map information at multiple scales, but this approach does not take into account the complementary role that contextual information in feature maps plays on semantics. In order for AMCC to make full use of contextual information and improve its representation ability, this study proposes a residual attention function fusion method to fuse contextual feature map information at different scales to improve the representation ability of the feature map, and further proposes a spatial attention mechanism for global pixel convolution responses.

5. Conclusion

In this paper, we implement a hardware system for real-time detection of human images. We checked the software implementation of the algorithm to ensure its correctness. Proposed Method: To match the found images, we use the Attention Mechanism Channel Cascade (AMCC) algorithm. Therefore, we capture surveillance video from the client and perform corner detection on the video. If an image edge is detected in a particular frame, the image edge is determined as a corner point to be identified. The extracted images are extracted and matched with images stored on the server. If a match is found, the client receives information about that image. Our results detect images and ensure that the detected image points indicate the server time and image matching accuracy. In future research, we will achieve image tracking within a specified frame, enabling highly accurate positioning and high-speed matching of surveillance camera positions. In order to improve the detection performance of small targets, researchers have conducted a lot of research in network structure, training strategies, data processing, etc., but there is still a large gap in the detection performance of small targets compared with the detection performance of large and medium-sized targets. Target size is one of the important factors that affect target detection performance. Currently, the detection accuracy of small targets is much lower than that of large and medium-sized targets in both public datasets and real images, resulting in frequent missed detections and false positives. However, small object detection has important applications in many natural scenes.

References

1. M. Muzammul, A. Algarni, Y. Y. Ghadi and M. Assam, "Enhancing UAV Aerial Image Analysis: Integrating Advanced SAHI Techniques With Real-Time Detection Models on the VisDrone Dataset," in IEEE Access, vol. 12, pp. 21621-21633, 2024, doi: 10.1109/ACCESS.2024.3363413.
2. W. Xiong, Z. Xiong and Y. Cui, "A Confounder-Free Fusion Network for Aerial Image Scene Feature Representation," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 5440-5454, 2022, doi: 10.1109/JSTARS.2022.3189052.
3. Y. Zhang, X. Gao, Q. Duan, L. Yuan and X. Gao, "DHT: Deformable Hybrid Transformer for Aerial Image Segmentation," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 6518805, doi: 10.1109/LGRS.2022.3222916.
4. M. Li, L. Zhao, N. Wang, X. Song and X. Wei, "Patch Grid-Based Quality Assessment for Aerial Visible-to-Infrared Image Translation," in IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, 2025, Art no. 8000105, doi: 10.1109/LGRS.2024.3519618.
5. B. R. A. Jaimes, J. P. K. Ferreira and C. L. Castro, "Unsupervised Semantic Segmentation of Aerial Images With Application to UAV Localization," in IEEE Geoscience and Remote Sensing Letters, vol. 19, pp. 1-5, 2022, Art no. 8020405, doi: 10.1109/LGRS.2021.3113878.
6. N. D. Vo et al., "Aerial Data Exploration: An in-Depth Study From Horizontal to Oriented Viewpoint," in IEEE Access, vol. 12, pp. 37799-37824, 2024, doi: 10.1109/ACCESS.2024.3371514.
7. S. Muhammad Jiskani, T. Hussain, A. Ali Sahito, F. Shaikh and L. Kumar, "Improving Insulators Detection Accuracy via Image Enhancement Techniques: Case of Indigenous Aerial Image Dataset," in IEEE Access, vol. 12, pp. 145582-145589, 2024, doi: 10.1109/ACCESS.2024.3474255.
8. S. K. Jangir and R. Bahmanyar, "Enhancing Very High-Resolution Satellite Images At 15 cm: A Novel Pipeline With Synthetic Data and Single

Image Superresolution," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 17, pp. 2670-2689, 2024, doi: 10.1109/JSTARS.2023.3346181.

9. H. Mizushina, K. Yamamoto and S. Suyama, "Unstable Depth Perception of Aerial Images in Crossed Mirror Array Can Be Controlled by Changing Fixation Distance," in IEEE Transactions on Industry Applications, vol. 58, no. 5, pp. 6793-6800, Sept.-Oct. 2022, doi: 10.1109/TIA.2022.3189971.
10. Y. Hua, L. Mou, P. Jin and X. X. Zhu, "MultiScene: A Large-Scale Dataset and Benchmark for Multiscene Recognition in Single Aerial Images," in IEEE Transactions on Geoscience and Remote Sensing, vol. 60, pp. 1-13, 2022, Art no. 5610213, doi: 10.1109/TGRS.2021.3110314.
11. H. Gao, Y. Yan, Y. He, J. Zhou, Z. Zhang and Y. Yang, "CAIL: Cross-Modal Vehicle Reidentification in Aerial Images Using the Centroid-Aligned Implicit Learning Network," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 18, pp. 2577-2588, 2025, doi: 10.1109/JSTARS.2024.3512579.
12. K. Zhao, S. Zeng, L. Zhou, T. Nie and S. Hao, "Prime Label Learning From Multilabel Aerial Image: A Novel Weakly Supervised Task," in IEEE Geoscience and Remote Sensing Letters, vol. 21, pp. 1-5, 2024, Art no. 6009405, doi: 10.1109/LGRS.2024.3401088.
13. T. K. Behera, S. Bakshi, M. Nappi and P. K. Sa, "Superpixel-Based Multiscale CNN Approach Toward Multiclass Object Segmentation From UAV-Captured Aerial Images," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 1771-1784, 2023, doi: 10.1109/JSTARS.2023.3239119.
14. Z. Wang, Z. Fu and J. Xu, "Efficient Superpixel-Based Seamline Detection for Large-Scale Image Stitching," in IEEE Geoscience and Remote Sensing Letters, vol. 22, pp. 1-5, 2025, Art no. 6005305, doi: 10.1109/LGRS.2025.3548266.
15. Z. Han, S. Zhang, Y. Su, X. Chen and S. Mei, "DR-AVIT: Toward Diverse and Realistic Aerial Visible-to-Infrared Image Translation," in IEEE Transactions on Geoscience and Remote Sensing, vol. 62, pp. 1-13, 2024, Art no. 5004213, doi: 10.1109/TGRS.2024.3405989.