Research Article

The Cognitive Transformation of Japanese Language Education by Artificial Intelligence Technology in the Wireless Network Environment

Su Zhang

School of Foreign Languages and International Education, Dalian Ocean University, Dalian 116023, Liaoning, China

Correspondence should be addressed to Su Zhang; zhangsu@dlou.edu.cn

Received 24 April 2022; Accepted 9 June 2022; Published 7 July 2022

Abstract

This study aims to solve the multiscale problems faced by the current classroom student behavior target detection based on the convolutional neural network (CNN) in the wireless network environment. Firstly, the recent reform of Japanese language education is introduced. Secondly, the multiscale problem research of classroom student behavior target detection is discussed. A CNN-based new extraction network is designed based on dilated convolution and pyramid features. An anchor reconstruction algorithm based on improved K-means clustering is presented for the self-made student behavior dataset. Finally, the performance of the designed algorithm is tested. The anchor reconstruction algorithm’s mean average precision is 83.2%, and the average intersection over union is 73.7%. The experimental results of this scheme outperform the original single-shot multibox detector and K-means algorithms. Compared with other algorithms, the designed multiscale detection algorithm of classroom student behavior has the best detection effect on Pascal visual object classes (VOC) dataset. The detection accuracy of the entire dataset is 79.8%. Overall, the multiscale detection algorithm for classroom student behavior has a better detection effect on the Pascal VOC dataset and has good generalization ability and robustness. This research can guide students to recognize their class status and make corresponding adjustments to improve their learning efficiency, which has essential research significance and application value.

1. Introduction

In recent years, with the construction of intelligent campuses in critical universities across the country, the smart campus combining teaching, scientific research, management, and the Internet of things (IoT) has been created in the wireless Internet environment, improving the efficiency of campus management and teaching. It is an important topic to change the thinking of Japanese teachers and continuously improve teaching methods through artificial intelligence (AI) empowerment in the AI era [1]. As a powerful auxiliary means of educational informationization, computer vision technology such as object detection is helpful to the improvement of classroom teaching evaluation and teaching management. It is one of the most important research directions in the future.

Students’ classroom behavior can fully reflect their learning effect and participation in teaching. Teachers can fully understand the learning effect of students and adjust teaching content and teaching methods through student behavior to improve the teaching quality [2, 3]. Target detection is mainly divided into traditional and CNN-based target detection algorithms [4]. Jiang et al. [5] proposed a multitask target detection model based on CNN. The model processes radar echo data directly, eliminating the need for time-consuming radar signal processing. The detection method can detect and locate targets in the multidimensional space of distance, velocity, azimuth, and elevation by using both time and frequency information. The experimental results showed that, compared with traditional radar signal processing methods and other advanced methods, the CNN-based detector has better detection performance and measurement accuracy in terms of distance, velocity, azimuth, and elevation and has stronger sensitivity to noise. Bai et al. [6] proposed an improved single-shot multibox detector (SSD) model based on deep feature fusion. The SSD algorithm adopted deep feature fusion between the object
2. Materials and Methods

2.2. Research on the Multiscale Problem of Classroom Student Behavior Target Detection. Object detection models based on CNNs are classified into the one-stage type based on region proposal and the two-stage type based on regression [10, 11]. The one-stage type based on region proposal directly uses the forward neural network to predict the prediction box of interest. The regression-based two-stage type first generates proposal boxes and then refines the regional features extracted from deep neural networks. However, both methods face the multiscale problem. Changes in the size of the object detection instance will affect the final effect of the detection. Multiscale image pyramids are an intuitive way to solve multiscale problems. However, the image pyramid method increases inference time and is therefore not suitable for real-time applications. The image of each layer of the Gaussian pyramid is used to subtract the predicted image after upsampling and Gaussian convolution of the image of the previous layer, generating a series of different images, that is, decomposed images. In the classroom student behavior target detection scene, cameras are set on both sides of the classroom blackboard to record the classroom situation of the students and then select clear still picture frames as the data set [12]. Because the distance between each row of students and the camera is different, the size of the students in the front row is large, and the size of the students in the back row is negligible. When the classroom is large, and the number of students is large, the size of the students in the front row is 3–5 times that of the students in the last row. In addition, the current object detection algorithms are all detected on Pascal visual object classes (VOC) dataset or the Common Objects in Context (COCO) dataset. Both datasets involve objects in natural scenes. If a series of regional candidate frames are established according to the size of the natural scene target to detect the student behavior target, it will inevitably lead to the candidate frame not matching the actual target, resulting in a further reduction of the
robustness of the model. Ultimately, this results in suboptimal object detection for classroom students.

2.3. Design of Network Structure

2.3.1. Network structure. The multiscale target detection network for classroom learning behavior includes a basic feature extraction network and high-level symmetric feature pyramid networks (FPN). The structure of the basic feature extraction network is shown in Figure 1. The first five groups of convolutional layers in the visual geometry group (VGG)-16 network are sequentially inputted into classroom learning images. Then, fully connected layers 6 and 7 are replaced with corresponding convolutional layers. The fully connected layer eight and the softmax classification layer are discarded. Finally, the remaining network adds four additional groups of convolutional layers at the end as the basic feature extraction network to realize the transformation of student behavior goals from the image domain to the feature domain.

In Figure 1, the students in the front row of the classroom and the students in the back row are of different sizes. This study designs a multiscale symmetric FPN including both top-down and bottom-up, as shown in Figure 2. The network generates the position information of the prediction box and the corresponding classification probability of the student behavior target. The original FPN structure only transfers the information of the highest-level feature map to the lower-level feature map, and the highest-level feature map does not add information about the low-level feature map [13, 14]. After a high-level symmetric FPN detection network, each predicted feature map contains the other five-dimensional feature maps' detailed information and semantic information.

2.3.2. Design of Multiscale Dilated Convolutional Layers. Dilated convolution can increase the size of the convolution kernel and increase the receptive field without changing the weight of the original convolution kernel. Dilated convolutions have been extensively used in semantic segmentation to fuse large-scale contextual information through dilated convolutions [15, 16]. The feature maps generated by dilated convolutions with different dilation rates are parallelized so that the neurons in the output feature map contain receptive fields of multiple scales, thereby improving the detection effect of classroom students’ behavioral targets. In addition, the propagation of gradients in multisized receptive fields tends to diverge, making it difficult for the network to converge. Therefore, a residual unit is added to solve this problem. Figure 3 reveals the specific network structure.

The principle of the residual unit is shown in the following equation:

$$x_L = x_l + \sum_{i=1}^{L-1} F(x_i, w_i),$$

where $x_L$ represents the $L$th layer output of the residual unit; $x_l$ represents the $l$th layer input of the residual unit; $x_i$ represents the $i$th feature layer; $w_i$ represents the...
corresponding convolution weight of the \( i \)th feature layer; \( F(x_i, w_i) \) represents the convolution operation on \( x_i \) and \( w_i \). The information of different deep networks can be fully integrated through the shortcut connection to meet the accuracy requirements of classroom student behavior target detection.

### 2.4. Region Candidate Box Reconstruction Algorithm

#### 2.4.1. Anchor Mechanism

The regional candidate frame generation mechanism can also be called the anchor mechanism, generating regional candidate frames. The generation of region candidate boxes directly affects the direction of target feature learning [17]. The introduction of the anchor mechanism reduces the repeated calculation of features and improves the detection speed and accuracy to a certain extent. However, the real-time performance of the algorithm is relatively poor due to a large number of calculations. SSD is a real high-quality, high-precision real-time detection algorithm. The SSD algorithm predicts candidate boxes on feature maps with different resolutions, directly divides \( S \times S \) grids according to their relative proportions to the original image, and generates candidate boxes of increasing sizes in the center of each grid [18, 19]. After the SSD algorithm obtains the region candidate frame, the probability of objects contained in these boxes is predicted, and the position offset of the box is calculated simultaneously. Then, the nonmaximum suppression algorithm filters out the candidate frame most likely to contain the target and readjusts the candidate frame in combination with the position offset of the frame to obtain the final detection result [20, 21].

#### 2.4.2. Analysis of the Scale of Classroom Student Behavior Datasets

Selecting the appropriate size of candidate boxes for the classroom behavior dataset and maximizing the dispersion of candidate boxes of different scales and widths to comprehensively cover the dataset can improve object detection results' effectiveness. Each candidate box size can be regarded as a cluster center of a small unit. Each small unit is a network optimization unit. Therefore, the classification and regression prediction parameters are optimized according to the corresponding training set. The core idea of the traditional \( K \)-means clustering algorithm is for a given data sample. The distance is used as the classification standard to divide the data within a short distance into a cluster. The number of clusters \( K \) can be set manually according to different needs [22, 23]. The specific process of the \( K \)-means clustering algorithm is as follows.

First, \( k \) cluster centers are randomly selected from the sample dataset \( X \), defined as set \( C \), as shown in the following equation:

\[
C = \{c_1, c_2, \ldots, c_k\},
\]

where \( 1 < k \leq n \), and \( n \) represents \( n \) objects.

Second, the Euclidean distance of all sample data to each cluster center is calculated. Then, the samples are divided into clusters corresponding to the nearest cluster centers [24]. Equation (3) describes the Euclidean distance.

\[
d(x_i, c_j) = \sqrt{\sum_{q=1}^{m} (x_{iq} - c_{jq})^2},
\]

where \( x_i \) represents the \( i \)th sample; \( c_j \) represents the \( j \)th cluster center; \( d(x_i, c_j) \) represents the Euclidean distance from the \( i \)th sample to the \( j \)th cluster center.

Third, the center of each cluster is calculated according to the centroid method. Equation (4) defines the \( j \)th cluster center \( c_j \).

\[
c_j = \frac{\sum_{x_i \in S_j} x_i}{|S_j|},
\]

where \( S_j \) represents the number of objects in the \( j \)th cluster.
Fourth, the last $k$ cluster centers are compared to determine whether the center has changed. If the center changes, skip to the second step for the next iteration; otherwise, the algorithm converges and outputs $k$ clusters.

2.4.3. Anchor Point Reconstruction Algorithm Based on Improved K-Means Clustering. The anchor reconstruction algorithm uses the width $w$ and height $h$ of all ground-truth boxes as input samples in the classroom behavior dataset. According to the distance metric, all the width and height samples are divided into categories corresponding to their nearest shape clustering centers. In this way, the distance between target candidate frames with significant shape differences is as considerable as possible, and the distance between target candidate frames with a similar shape is as tiny as possible. Here, a new similarity distance metric, the intersection distance metric over the union, is employed, as shown in the following equation:

$$d(\text{Other, Center}) = 1 - \text{IoU}(\text{Other, Center}),$$

where $d(\text{Other, Center})$ denotes the distance between other borders and the cluster center border; $\text{IoU}(\text{Other, Center})$ represents the intersection ratio of other borders and the cluster center border, as shown in the following equation:

$$\text{IoU}(\text{Other, Center}) = \frac{S(\text{Other} \cap \text{Center})}{S(\text{Other} \cup \text{Center})}.$$  

where $S(\text{Other} \cap \text{Center})$ denotes the area of the intersection of the two borders; $S(\text{Other} \cup \text{Center})$ signifies the area of the union of the two borders.

All the width and height samples are randomly selected from the $k$ cluster center points and performed initial division. These central stores are then iteratively optimized until width and height clustering results are obtained for all student behavioral targets. The genetic algorithm is used for the initial center point of the global optimization algorithm [25]. The fitness function in the genetic algorithm is a crucial part of the anchor reconstruction algorithm. Each anchor box in the study is a training sample. Each anchor box is either marked as a background or associated with an actual edge box, generating many anchor boxes and producing many negative class samples. Equation (7) describes the fitness function designed here.

$$f = \frac{D_{\text{between}}(d(\text{Center}, \text{Center}_j))}{1 + D_{\text{in}}(d(x_i^{(j)}, \text{Center}_j))},$$

$$D_{\text{between}}(d(\text{Center}, \text{Center}_j)) = \frac{2}{k(k-1)} \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} d(\text{Center}, \text{Center}_j),$$

$$D_{\text{in}}(d(x_i^{(j)}, \text{Center}_j)) = \frac{1}{k} \sum_{j=1}^{n_i} \sum_{i=1}^{n_j} \frac{d(x_i^{(j)}, \text{Center}_j)}{n_j},$$

Figure 3: Cascaded multidilation rate convolution structure with residual units.
where \( d(Center_i, Center_j) \) represents the intersection ratio distance between the \( i \)th and the \( j \)th cluster center; \( d(x^{ij}_i, Center_j) \) represents the intersection ratio distance between the \( i \)th sample of the \( j \)th class and the \( j \)th cluster center; \( n_j \) represents the \( j \)th class total number of samples.

Moreover, the classification and offset of each anchor box are labeled. Suppose an anchor box A is assigned a ground-truth bounding box B. On the one hand, the class of anchor box A will be marked the same as B. On the other hand, the offset of anchor box A will be marked according to the relative position of the center coordinates of B and A and the relative size of these two boxes. Given the different positions and sizes of different boxes within a dataset, the model can apply a transformation to those relative positions and dimensions to get an evenly distributed, easily adaptable offset. Figure 4 displays the flow of the anchor reconstruction algorithm designed here.

2.5. Experimental Configuration. This experiment trains and validates the multiscale detection method of classroom student behavior based on a deep learning framework and makes a data set of student behavior in class for model verification and evaluation. The study tests five thousand test images. In addition, the generalization ability of the model has been verified on the public Pascal VOC dataset. The 2,501 images in Pascal VOC2007 and 5,717 in Pascal VOC2012 are used as the training set, and the 4,952 images in Pascal Voc2007 are used as the test set. The specific environment configuration is a central processing unit Intel Core i7-6700 and 4 Graphic Processing Units. The model number is GTX 1080Ti. The operating system is Windows 10, and the programming language is Python 2.7, Matlab 2014b.

In addition, the slope of the LeakyReLU activation function in the detection network is set to 0.2; the momentum is 0.9; the batch size is 16; the times of training are 90; the initial learning rate is 0.001. After 50 iterations, the learning rate declines to \( 10^{-4} \) by a ratio of 0.1. At 70 iterations, it is reduced to \( 10^{-5} \) by a ratio of 0.1. Throughout the training process, the model learning rate decays to 0.0005. Average precision (AP) is used to measure object detection accuracy. Frames per second are a measure of the speed of object detection. The average intersection of the models is used to measure how well the predicted box matches the ground truth. The parameter Mb is used to measure the space complexity of the model.

3. Results and Discussion

3.1. Comparison of Data Enhancement Effects. The diversity and complexity of training samples play an essential role in the improvement of the model detection effect and the improvement of the generalization ability of classroom students’ behavioral goals. Here, data augmentation methods are used to improve the overall performance of model training. Four data augmentation methods, flipping, cropping, rotating, and adding Gaussian noise, are analyzed to generate new sample numbers. The number of samples is one and three times the initial training set for comparative experiments. The comparison results of the four data enhancement techniques are shown in Figure 5.

As can be seen from Figure 5, data expansion and enhancement techniques are mainly divided into cropping, rotation, and Gaussian noise filtering. Among them, the three methods of flipping, cropping, and rotating can improve the accuracy of the trained model. Before the cropping method is adopted, the recognition accuracy of the data enhancement technology is only 77.5%. After the double enhancement of the cropping method, the recognition accuracy of the data expansion technology can be improved to 79.8%; the recognition accuracy of the data expansion can be improved to 81.5% under the cropping method after the triple enhancement. In addition, the model recognition accuracy is 77.8% before using flip technology to enhance the data expansion effect; after double-enhancing using flipping technology, the model recognition accuracy reaches 81.8%; the model recognition accuracy after triple-enhancing attains 83%. The higher the training samples, the better the effect. In addition, Gaussian noise will reduce the recognition effect of the trained model to a certain extent. Therefore, the combined effects of flipping, cropping, and rotating methods are further tested for comparison, as presented in Figure 6.

In Figure 6, the hybrid method of flipping, cropping, and rotating has the best effect. When the amount of data is doubled, the accuracy rate is 83.5%; when the amount of data is tripled, the accuracy rate is 84.9%. Therefore, in all experiments, the training data are in the form of a combination of three methods of flipping, cropping, and rotating, and the data set is tripled to obtain the best detection effect for classroom student behavior.

3.2. Comparison of Anchor Clustering Results. The original SSD algorithm, the traditional K-means algorithm, and the anchor reconstruction algorithm are compared. Figure 7 presents the comparison results of the three algorithms.
In Figure 7, the MAP of the designed anchor reconstruction algorithm is 83.2%, and the AIoU is 73.7%, which is better than the original SSD algorithm and K-means algorithm in both MAP and AIoU. The anchor reconstruction algorithm can overcome the sensitivity of the K-means algorithm to the initial cluster centers. The shape of the predicted candidate box represented by the clustering results can better fit the real shape of the target on the classroom student behavior dataset. The robustness has been dramatically improved compared with the traditional K-means and the original SSD algorithms.

Performance analysis of multiscale detection algorithm of student behavior in the classroom.

The SSD, SSD+FPN, YOLOv2, and faster R–CNN algorithms are selected for inspection and comparison with the proposed multiscale detection algorithm for classroom student behavior. The specific detection results are shown in Figure 8.

Figure 8 suggests that the multiscale detection algorithm designed for classroom student behavior is significantly better than other algorithms in terms of all-around performance. The detection speed of the multiscale detection algorithm of student classroom behavior is not much different from that of the SSD algorithm and SSD+FPN. However, this algorithm has an accuracy of 85.1%, higher than other algorithms. This is because it adopts parallel dilated convolution and symmetric FPN, enabling the network to fully integrate the information of different depth networks, which is conducive to detecting student behavioral targets. In addition, the model using the faster R–CNN

![Figure 5: Comparison of the effects of four data augmentation techniques.](image1)

![Figure 6: Comparison of combined effects of flip, crop, and rotate methods.](image2)
algorithm has the highest system data transfer rate, far exceeding the other several detection models.

The comparison of the detection results of the five detection algorithms for the five types of actions of writing, listening, raising hands, sleeping, and answering questions is shown in Figure 9.

In Figure 9, the AP values of the multiscale detection algorithm reported here are 84.4% for writing, 85.6% for listening, 87% for raising a hand, 79.8% for sleeping, and 79.8% for answering. The detection AP value of the problem is 88.5%. The detection effects of listening, raising a hand, sleeping, and answering questions are better than other algorithms. From the overall trend, since raising a hand has prominent behavioral characteristics, the detection effect of this behavior is relatively good. On the contrary, the sleeping behavior is easily blocked by other behaviors, so the detection effect is poor.

The algorithm proposed here is trained and tested on the public Pascal VOC dataset to verify its generalization ability and effectiveness. Figure 10 presents the detection MAP results of the five algorithms on the Pascal VOC dataset.

In Figure 10, compared with other algorithms, the multiscale detection algorithm of classroom student behavior has the best detection effect on the Pascal VOC dataset.
dataset. The detection accuracy of the entire dataset is 79.8%. Overall, the multiscale detection algorithm for classroom student behavior shows excellent generalization ability and robustness.

To sum up, the recognition effects of students’ classroom behavior of the algorithm proposed here are better than other algorithms. The average recognition accuracy of several algorithms is analyzed. The AP value of the algorithm proposed here is 84.4% for writing, 85.6% for listening, 87% for raising hands, 79.8% for sleeping, and 79.8% for answering. On the Pascal VOC dataset, the detection accuracy of the algorithm is 79.8%. Overall, the multiscale detection algorithm for classroom student behavior is superior to the other algorithms regarding detection effect and robustness.

4. Conclusions
The detection of student behavior in the classroom environment combines classification tasks and positioning tasks, reflecting students’ participation in classroom teaching and playing a vital role in promoting Japanese language education and teaching informatization. Aiming at the multiscale problem in which the size of the students in the front row and the back row is quite different in the classroom scene, the
network structure and the reconstruction of the candidate frame size are mainly studied. A multiscale detection network is designed based on the SSD algorithm for the network structure under the multiscale problem. The FPN with a symmetric system is designed for the shortcoming of the lack of shallow detail information in the deep feature map of FPN. In addition, a candidate box reconstruction algorithm based on improved K-means clustering is designed to obtain the width and height dimensions of the true distribution of the dataset. The research results demonstrate that the anchor reconstruction algorithm has greater robustness than the traditional K-means algorithm and the original SSD algorithm. The algorithm is superior to other algorithms in the comprehensive performance of the multiscale detection algorithm of student behavior in the classroom and has good generalization ability and robustness. The disadvantage is that the designed classroom student behavior detection algorithm does not have an advantage in the number of model parameters. Therefore, the detection algorithm can be further optimized from the perspective of model compression and network lightweight in the future.

Data Availability

The data underlying the results presented in the study are available within the manuscript.

Conflicts of Interest

The authors declare that there are no potential conflicts of interest.

References

[1] J. A. Ruipérez-Valiente, D. Jaramillo-Morillo, S. Joksimović, V. Kovanošić, P. J. Muñoz-Merino, and D. Gašević, “Data-driven detection and characterization of communities of accounts collaborating in MOOCs,” *Future Generation Computer Systems*, vol. 125, pp. 590–603, 2021.

[2] S. Li, S. P. Lajoie, J. Zheng, H. Wu, and H. Cheng, “Automated detection of cognitive engagement to inform the art of staying engaged in problem-solving,” *Computers & Education*, vol. 163, Article ID 104114, 2021.

[3] C. K. Chiu and J. C. Tseng, “A bayesian classification network-based learning status management system in an intelligent classroom,” *Educational Technology & Society*, vol. 24, no. 3, pp. 256–267, 2021.

[4] L. Huang, Q. Fu, M. He, D. Jiang, and Z. Hao, “Detection algorithm of safety helmet wearing based on deep learning,” *Concurrency and Computation: Practice and Experience*, vol. 33, no. 13, Article ID e6234, 2021.

[5] W. Jiang, Y. Ren, Y. Liu, and J. Leng, “A method of radar target detection based on convolutional neural network,” *Neural Computing & Applications*, vol. 33, no. 16, pp. 9835–9847, 2021.

[6] D. Bai, Y. Sun, B. Tao et al., “Improved single shot multibox detector target detection method based on deep feature fusion,” *Concurrency and Computation: Practice and Experience*, vol. 34, no. 4, Article ID e6614, 2022.

[7] H. Wei, “Cultivation model for autonomous learning ability of Japanese majors,” *International Journal of Emerging Technologies in Learning (iJET)*, vol. 16, no. 04, p. 201, 2021.

[8] S. H. S. Al Hasani and N. A. Husin, “A review of digital transformation of education in Oman,” *Journal of Industrial and Production Engineering*, vol. 39, no. 2, pp. 41–59, 2021.

[9] H. Komatsu and J. Rappleye, “Reimagining modern education: contributions from modern Japanese philosophy and practice?” *ECNU Review of Education*, vol. 3, no. 1, pp. 20–45, 2020.

[10] F.-P. An, J.-E. Liu, and L. Bai, “Object recognition algorithm based on optimized nonlinear activation function-global convolutional neural network,” *The Visual Computer*, vol. 38, no. 2, pp. 541–553, 2022.

[11] G. Hu, K. Wang, and L. Liu, “Underwater acoustic target recognition based on depthwise separable convolutional neural networks,” *Sensors*, vol. 21, no. 4, p. 1429, 2021.

[12] B. Sun, Y. Wu, K. Zhao et al., “Student Class Behavior Dataset: a video dataset for recognizing, detecting, and captioning students’ behaviors in classroom scenes,” *Neural Computing & Applications*, vol. 33, no. 14, pp. 8335–8354, 2021.

[13] T. Zhang, X. Zhang, and X. Ke, “Quad-FPN: a novel quad feature pyramid network for sar ship detection,” *Remote Sensing*, vol. 13, no. 14, p. 2771, 2021.

[14] X. Li, Z. Xu, X. Shen, Y. Zhou, B. Xiao, and T.-Q. Li, “Detection of cervical cancer cells in whole slide images using deformable and global context aware faster RCNN-FPN,” *Current Oncology*, vol. 28, no. 5, pp. 3585–3601, 2021.

[15] A. Das, K. Bhardwaj, and S. Patra, “Deep convolution neural network with automatic attribute profiles for hyperspectral image classification,” *Multimedia Tools and Applications*, vol. 80, no. 28–29, pp. 35365–35385, 2021.

[16] G. Li, V. L. Knoop, and H. van Lint, “Multistep traffic forecasting by dynamic graph convolution: interpretations of real-time spatial correlations,” *Transportation Research Part C: Emerging Technologies*, vol. 128, Article ID 103185, 2021.

[17] Y. Cong, Y. Feng, H. Ni et al., “Anchor-locker binding mechanism of the coronavirus spike protein to human ACE2: insights from computational analysis,” *Journal of Chemical Information and Modeling*, vol. 61, no. 7, pp. 3529–3542, 2021.

[18] S. Zhou and J. Qu, “Enhanced SSD with interactive multiscale attention features for object detection,” *Multimedia Tools and Applications*, vol. 80, no. 8, pp. 11539–11556, 2021.

[19] G. Yadgar, M. Gabel, S. Jaffer, and B. Schroeder, “SSD-based workload characteristics and their performance implications,” *ACM Transactions on Storage*, vol. 17, no. 1, pp. 1–26, 2021.

[20] I. Ahmed, M. Ahmad, A. Ahmad, and G. Leon, “IoT-based crowd monitoring system: using SSD with transfer learning,” *Computers & Electrical Engineering*, vol. 93, 2021.

[21] C. Sun, Y. Ai, S. Wang, and W. Zhang, “Mask-guided SSD for small-object detection,” *Applied Intelligence*, vol. 51, no. 6, pp. 3311–3322, 2021.

[22] J. Estupiñán Ricardo, J. J. Domínguez Menéndez, and I. F. Barcos Arias, “Neuroptic K-means for the analysis of earthquake data in Ecuador,” *Neutrosophic Sets and Systems*, vol. 44, no. 1, p. 29, 2021.

[23] Z. Zhang, Q. Feng, J. Huang, Y. Guo, J. Xu, and J. Wang, “A local search algorithm for k-means with outliers,” *Neurocomputing*, vol. 450, pp. 230–241, 2021.

[24] E. Zhang, H. Li, Y. Huang, S. Hong, L. Zhao, and C. Ji, “Practical multi-party private collaborative k-means clustering,” *Neurocomputing*, vol. 467, pp. 256–265, 2022.

[25] S. Fatemi-Anaraki, M. Mohktarzadeh, M. Rabbani, and D. Abdolhamidi, “A hybrid of K-means and genetic algorithm to solve a bi-objective green delivery and pick-up problem,” *Journal of Industrial and Production Engineering*, vol. 39, no. 2, pp. 146–157, 2022.