ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) https://doi.org/10.17559/TV-20230926000963
Original scientific paper

# Visual Communication Design of Weak and Small Target Images Based on Image Processing Model and Data Fusion

Xu ZHANG

**Abstract:** The essence of computer vision technology is to combine computers with image data to enhance their understanding and perception abilities. One of the major research hotspots in computer vision technology is the object recognition, and in the field of object recognition, a major challenge is the recognition of weak and small target images. In response to the current difficulty in detecting weak and small target images, a visual communication design with image processing models and data fusion was proposed. The Single Shot MultiBox Detector was improved to construct a weak and small target detection model, and the feature fusion method was combined to enhance its weak and small target detection ability. The ablation experiment showed that in contrast with the original model, the improved model improved the detection ability of weak and small targets by 26.54%, and the overall accuracy improved by 11.05%. Two other advanced algorithms were selected for comparison with the research algorithm, and the accuracy of the research algorithm was better, with a higher accuracy of 47.67% -79.56% than the comparison algorithm. The response time was shorter, reaching 0.62 seconds. Visual communication time and success rate performed better, with a communication time lead of 9 s to 19 s and a success rate lead of 16% to 27%. In summary, the algorithm proposed by the research institute has higher accuracy in detecting small and weak images, as well as higher visual communication efficiency and success rate, which has stronger practical significance in the field of computer vision.

**Keywords:** attention mechanism; detection of weak and small targets; feature fusion; SSD algorithm

## 1 INTRODUCTION

Computer vision technology is a process of simulating, analyzing, and understanding human vision using computer algorithms and methods. It covers many different tasks and application fields, including image and face recognition, object detection and tracking, scene understanding, etc. [2]. The field of target detection is the focus of research all along. In some visual detection fields, due to long distances or the small volume of the object itself, the object only presents a few pixels on the image, which becomes a weak target image. Dim target image detection is an important task in the field of computer vision, whose main goal is to detect and locate small objects accurately and effectively in a large number of images. These objects are often defined as "dim targets" because they have a relatively low signal-to-noise ratio (SNR), low contrast, or small size [3]. If it is necessary to recognize and detect such targets, visual communication must be used as a medium for information transmission [4]. However, this type of target exhibits low contrast and SNR, and image detection techniques for such weak targets need to be targeted to enhance their ability to detect weak and small target (WST) [5]. Therefore, research has proposed a WST image detection method based on image processing models and data fusion. Based on the Single Shot MultiBox Detector (SSD) algorithm, WST detection is enhanced to adapt to the research target. An image processing model is constructed, and feature fusion (FF) is used to enhance data information, with a slight increase in data volume as a prerequisite, which further enhances the model's detection ability for WSTs. There are two innovative points in the research. The first point is the design of a new structure for shallow features based on FF. The second point is to improve the context module and attention mechanism module with the goal of light weight. The study is broken into four parts. The first part is a summary of the field of computer image processing for primary weak target detection. The second and the third parts are the construction and the verification of the effectiveness of the proposed method, respectively. The fourth part is a summary and outlook for the study.

### 1.1 Related Works

Image processing models are computer algorithms or deep learning models used for processing and analyzing digital images. These models are widely used in many image processing tasks, including semantic segmentation, image generation, and style transfer. Wang J. proposed a stock prediction model based on image recognition technology, which addressed the issue of traditional mathematical models not being intuitive enough in financial analysis by using image normalization technology. The research outcomes demonstrated the effectiveness and practicality of the proposed method [6]. Navi K. et al. proposed a new approximate full adder for carbon nanotubes, which effectively improved the speed and stability of image monitoring [7]. Da Hai et al. summarized the existing image processing evaluation methods for material degradation through a review method, and made suggestions for the existing shortcomings and development direction [8]. Khaddam H. S. et al. put forward a novel method for confirming the diameter of cotton yarn using image processing technology combined with artificial neural networks. The research findings have proven the function and practicality of this method [9]. Lopez N. R. et al. proposed a method for analyzing the real-time state of pressure vessels through image processing technology, which could more effectively produce digital models of vessels [10].

WST image detection is a subdivision field in computer vision, which means the task of accurately detecting and locating small targets in the image. These targets may be small in size, low in contrast, or obscured by other objects, making them challenging for algorithms. Yang B. et al. raised a new detection algorithm based on YOLOv4, which used convolutional kernel modules of different sizes to improve the model's feature extraction ability. The experimental outcomes demonstrated that this

method improved the accuracy of small object detection [11]. Zhao M. et al. raised a detection model that combined frequency modulation and spectroscopy, fully utilizing spectral information and effectively improving the detection ability for weak targets [12]. Sun M. et al. proposed a road infrared target detection model with Efficient Net, which effectively reduced the noise in infrared images and improved the detection accuracy for small targets [13]. Zhang Ty et al. proposed a small object detection method with infrared search and tracking technology, which transformed the small object detection problem into an optimization problem. The experiment outcomes demonstrated the effectiveness and excellence of this method [14]. Zhang L. et al. proposed an infrared algorithm for small object detection, aiming to develop an efficient, robust, and reliable small object detection method. Experiment findings have demonstrated this method's performance [15].

One-stage object detection algorithm is a kind of object detection method in the field of computer vision. The main feature is that the object detection and classification can be completed through one forward propagation (single stage). Compared with the two-stage object detection algorithm, the one-stage object detection algorithm was faster, but it might sacrifice a little in accuracy. Xing Z. et al. proposed an efficient detection network search strategy to solve the problem of speed reduction caused by large number of parameters in the YOLOv4 model. The experimental results showed that this method was effective for model simplification and improves the accuracy of the model [16]. Under the background of Industry 4.0 reform, Yan J. et al. proposed a transfer learning network based on YOLOv3, in which VGG 16 was adopted to improve the original model. And the experimental results showed that it effectively realized automatic identification, monitoring and analysis of data [17]. Guo L. et al. proposed an accurate and fast SSD detection model to further improve the performance of target detection. Compared with the existing deep learning image recognition model, it can generate more detailed feature maps and has better performance [18]. Wang G et al. proposed a YOLO target detection network for colleges and universities. By improving its original framework, the accuracy rate and recall rate of lightweight target detection were effectively improved, and the computational complexity is reduced [19]. Aiming at the current low real-time performance and accuracy of pedestrian tracking, Yang S. et al. proposed a multi-stage target detection model combined with SSD algorithm, and the experimental results proved that the model had excellent real-time performance, high stability and adaptability [20].

FF refers to combining multiple features from different sources or different types to obtain a more comprehensive and rich feature representation. In many machine learning and deep learning tasks, FF plays an important role in improving model performance and robustness. Barbhuiya A. et al. proposed a gesture recognition model combining convolutional neural network and Zernike matrix to address the challenges faced by gesture recognition in the field of computer image recognition, and adopted FF to enhance the recognition efficiency of local images. Experimental results proved the accuracy and recognition accuracy of this method [21]. Song Y. et al. proposed an indoor human behavior recognition method based on Wi-Fi perception combined with FF to realize effective recognition of indoor human behavior, which had positive significance for indoor behavior monitoring of elderly people living alone [22]. To solve the problem of ultrasonic diagnosis of defects in austenitic stainless steel, Zhang R. et al. proposed a new defect diagnosis method based on multi-domain data FF. The experimental results showed that this method has comprehensively improved the performance of defect diagnosis and had a driving effect on ultrasonic missing element diagnosis [23]. Cui X. et al. proposed a classification model by fusing image features and depth features to improve the performance of the MRI aided diagnosis model. Experimental results showed that the accuracy rate of this method was as high as 98.66%, and it had higher performance [24]. Aiming at the challenges in crowd counting, Wang L. et al. proposed a multi-layer FF network framework for single-image crowd technology, and the experimental results proved the excellent performance of this method [25].

Based on the above content, WST image detection is currently an extremely important research direction in computer vision. However, the biggest challenge it currently faces is the extremely high detection difficulty of WSTs, which often stems from the extremely low contrast and SNR of the small targets themselves. Therefore, a visual communication method for WST images based on image processing models and data fusion has been proposed. The SSD algorithm is used as the basic algorithm to improve the targeted detection of WSTs, and combined with FF to enhance the detection ability of WSTs.

## 2 CONSTRUCTION OF A WST IMAGE DETECTION METHOD WITH IMPROVED SSD AND DATA FUSION

The study selects SSD algorithm as the basic algorithm, and targeted improvements are made by adding a dedicated prediction head for WSTs to enhance the detection ability of SSD algorithm for WSTs. To further enhance its ability to detect WSTs, combined with the idea of data fusion, an FF method is adopted to further enhance the features of WSTs, thereby enhancing the algorithm's ability to detect WSTs.

### 2.1 WST Image Detection Method with Improved SSD

The SSD algorithm is a relatively classic object detection algorithm. Its most significant advantage compared to other object detection algorithms is its extremely fast speed, while still keeping a high average accuracy [26]. The basic SSD network structure diagram is expressed in Fig. 1. The SSD algorithm improves the detection accuracy of WSTs by adopting a fully convolutional neural network structure and multi-layer FF method. It detects targets on feature maps of different scales by adding additional convolutional layers and feature map fusion techniques to capture targets of different sizes. This multi-scale design enables SSDs to simultaneously predict multiple target boxes of different sizes and corresponding category information in a forward propagation. In addition, SSD also uses a prior box mechanism to adapt to different target shapes and scales.

This design allows SSDs to adapt to the detection needs of various WSTs.

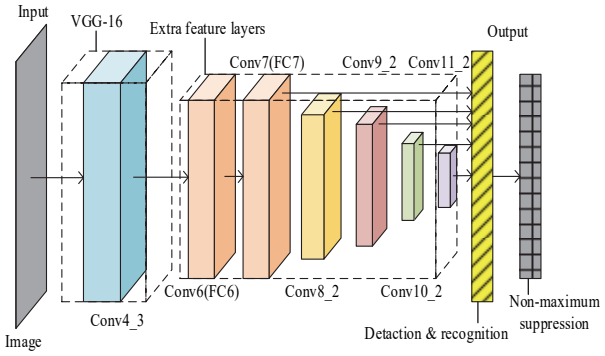


**Figure 1** Schematic diagram of network structure of SSD

In the original SSD network, when predicting targets, it is usually necessary to generate a series of default boxes in each feature map. It assumes that the *m*-th layer feature map is currently utilized for target detection; the size calculation of the default boxes in each layer feature map is shown in Eq. (1).

$$s_k = s_{\min} + \frac{s_{\max} - s_{\min}}{m-1}(k-1), k \in [1, \cdots, m] \tag{1}$$

In Eq. (1), $s_{\min}$ and $s_{\max}$ sever as the mini and the max ratio of the default boxes in the feature layer to the input image, respectively. Usually, its specific settings are shown in Eq. (2).

$$\begin{cases} s_{\min} = 0.2 \\ s_{\max} = 0.95 \end{cases} \tag{2}$$

Then, different aspect ratios ( $aspect\_ratio$ ) need to be used to generate multi-scale default boxes on each layer's feature map. The expression of the width and height of the multi-scale default boxes is expressed in Eq. (3).

$$\begin{cases} Width_k^a = s_k \sqrt{a_r} \\ Height_k^a = s_k / \sqrt{a_r} \end{cases}, k \in [1, \cdots, m], a \in a_r \tag{3}$$

When $aspect\_ratio = 1$ , it is necessary to further add a square default box, as shown in Eq. (4).

$$s_k^{'} = \sqrt{s_k s_{k+1}} \tag{4}$$

However, there is still room for improvement in SSD's monitoring of WSTs, and its performance in accuracy testing for WSTs is extremely poor. Therefore, it is necessary to improve and strengthen its ability to detect WSTs in a targeted manner. The original convolutional neural network design in SSD networks has shortened response time and effectively reduced the number of parameters. It has excellent performance in images with large spans, high contrast, and SNR, but often performs poorly in detecting WSTs. Therefore, the study chooses to improve the original SDD network by adding small target prediction heads to the existing prediction heads in the network for monitoring WSTs, which can better alleviate the accuracy impact brought by the target scale being too small, and can greatly increase the detection ability of SSD network for WSTs. The improved SSD network structure is shown in Fig. 2.

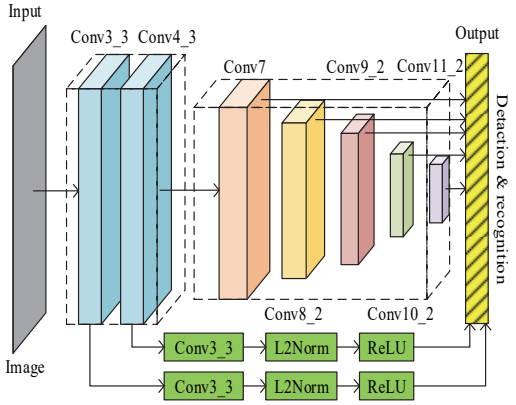


**Figure 2** Schematic of the SSD network structure with the new prediction head structure added

In the study, it improves the original $Conv3\_3$ feature layer in the network to a new feature layer targeting WSTs. The feature map size of the Conv3_3 feature layer is $75 \times 75 \times 512$ pixels, which has more details and is more conducive to monitoring WST image information. Further research has added convolution, normalization, and activation functions to the added WSTimage prediction head to maximize its WST image detection performance.

### 2.2 Construction of a WST Image Detection Method Based on Improved SSD and FF

Furthermore, to enhance the detection ability of SSD networks for WSTs, and to address the limited ability of each feature layer in the original SSD network, the shallow feature pyramid structure of SSD networks is redesigned. Using the idea of data fusion and the method of multi-scale FF, the feature information of WST image targets is enhanced [27]. Its essence is to use a feature pyramid structure to fully connect feature layers of different depths, overcoming the problem of the original network only using shallow feature maps, which is easily affected by image back noise. The improved shallow feature pyramid structure in the study is shown in Fig. 3. The original size of the Conv8_2 feature map in SSD networks is $10 \times 10 \times 512$ pixels, which may contain too much background noise due to its large range. So the study only uses Conv3_3,Conv4_3,Conv7 feature layer to reconstruct the pyramid structure, and further strengthens its feature extraction ability for Conv3_3,Conv4_3 feature layer.

The improved shallow pyramid structure of the research institute only uses three shallow feature layers and establishes a top-down pyramid structure based on the high-resolution features provided by them. By discarding deep feature information and fusing shallow feature information, shallow semantic information is fused while ensuring sufficient information, thereby improving the

network's feature detection ability for WSTs. There is still a problem of difficulty in locating WSTs in SSD networks.

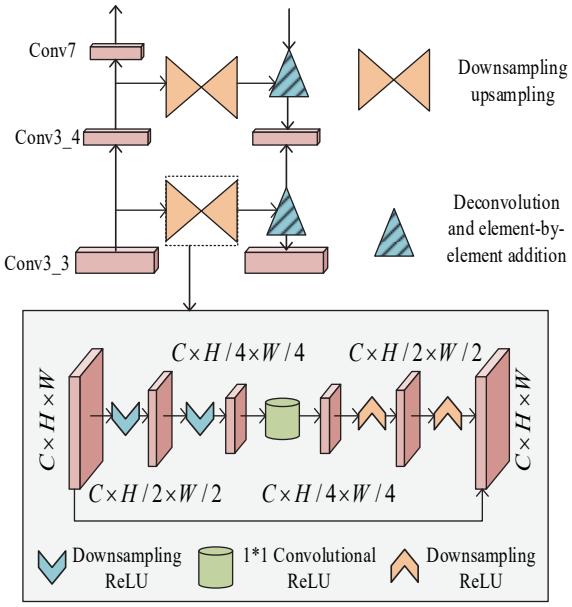


**Figure 3** Schematic diagram of shallow feature pyramid structure

Therefore, a multi-scale contextual FF module inspired by Inception networks has been proposed to fuse contextual features of different scales for detecting targets. In this module, three different contextual feature extraction partitions are used for feature dimensionality reduction, convolution, and extraction processing. To ensure that contextual information does not overwhelm the features of the detection target itself, the output channels of the two partitions are set to the original $\frac{1}{4}$, and the output channels of the remaining partition are set to $\frac{1}{2}$. Then it connects the three partitions and restores the original number of channels. Finally, the stacking method is used to fuse the features of the detection target itself and its contextual features. Its details are shown in Eq. (5).

$$y = x + C\left(F_{b1}(x) + F_{b2}(x) + F_{b3}(x)\right) \tag{5}$$

In Eq. (5), $y$ refers to the output result; $x$ means the input feature map; $C$ denotes the connection operations performed in the module; $F_{b1}, F_{b2}, F_{b3}$ indicate the three partitions in the module. Through this design, the target environment association information in weak target images can be more effectively extracted, providing more data for the detection of weak targets and enhancing the model's detection ability for weak target images. The structure of the proposed multi-scale context FF module is shown in Fig. 4.

Due to the small size or low contrast of weak targets in the image, their detection and localization are relatively difficult. The attention mechanism can help the network focus on key areas when processing images, thereby improving the detection performance of WSTs. Therefore, the study introduces attention mechanism and combines the attention module of spatial attention combined with channel attention with the multi-scale context FF module to construct an improved context attention FF module in the study. Its structure is shown in Fig. 5.

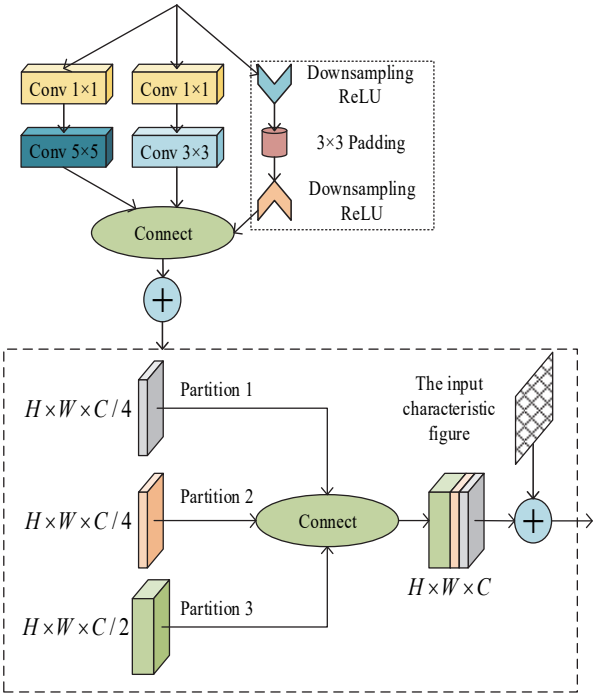


**Figure 4** Schematic diagram of the multi-scale context FF module structure

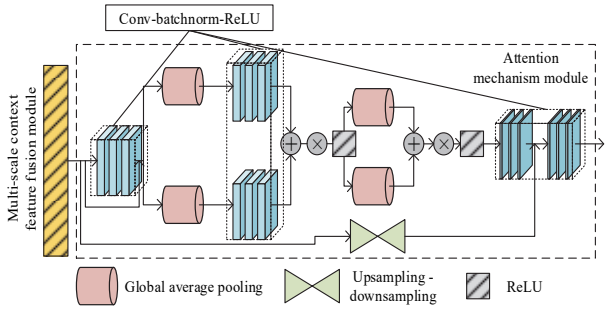


**Figure 5** Schematic diagram of the improved context attention FF module structure

After the multi-scale context FF module, it will concatenate convolution, normalization, and activation operations. It assumes that the original size of the input feature map is $H \times W \times C$. Convolution and normalization operations are performed through the maximum pooling and average pooling processing of the attention module, and addition operations are performed on each element. Finally, the attention image information of the channel with size $1 \times 1 \times C$ is output. It multiplies this information with the original feature map along the channel dimension element by element, and ultimately generates an image with channel attention features. The process can be expressed as shown in Eq. (6).

$$M_c(x) = \sigma\left(\left(Cnr\left(F^c_{\text{avg}}(x)\right) + Cnr\left(F^c_{\max}(x)\right)\right) \cdot x_c\right) \tag{6}$$

In Eq. (6), $Cnr$ severs as the convolution, normalization, and activation operations performed in the module; $\sigma$ means ReLU activation; $F^c_{\text{avg}}$ expresses global

average pooling; $F_{\max}^{c}$ refers to global maximum pooling, and $x_c$ denotes the channel dimension of the feature map. After the feature map passes the activation operation, further information extraction operations for spatial attention are required. The channel attention feature map first undergoes maximum pooling and average pooling processing, and outputs feature maps with a size of $H \times W \times 1$. Then, these two feature maps are further concatenated and dimensionally reduced to obtain spatial attention feature information with a size of $H \times W \times 1$. Finally, they are multiplied by each element to output the final feature. The process is represented by Eq. (7).

$$M_s(x) = \sigma\left(T\left(F_{\text{avg}}^{s}(x), F_{\max}^{s}(x)\right) \cdot x_{h,w}\right) \tag{7}$$

In Eq. (7), $\sigma$ denotes the ReLU activation function; $F_{\text{avg}}^{c}$ indicates the global average pooling operation; $F_{\max}^{c}$ means the global maximum pooling operation; $T$ represents the concatenation dimensionality reduction, and $x_{h,w}$ indicates the dimension of the feature map. For image detection of WSTs, the uneven distribution of WSTs can seriously affect the training process of the model, which in turn hinders the model's ability to detect WSTs. Therefore, the study chooses to use the Facal Loss function to propose a solution to the problem of sample non-uniformity [28]. The more classic loss function is the standard cross entropy, as indicated in Eq. (8).

$$L_{ce} = L(p, y) = \begin{cases} -\log(p), y = 1 \\ -\log(1-p), y = -1 \end{cases} \tag{8}$$

In Eq. (8), $p \in [0,1]$ expresses the probability distribution of a positive sample, where $y = 1$ or $y = -1$. For the convenience of representation, it further unifies $p$ and $1-p$ and sets $P_t$, and the specific representation is shown in Eq. (9).

$$P_t = \begin{cases} p, y = 1 \\ 1-p, y = -1 \end{cases} \tag{9}$$

Then the standard cross entropy loss function can be further represented as shown in Eq. (10).

$$L_{ce}(p, y) = L_{ce}(P_t) = -\log(P_t) \tag{10}$$

However, when faced with imbalanced samples, this function will tilt due to imbalanced samples, and massive negative samples will result in a decrease in the model's detection performance for samples of fewer categories. The usual improvement method is to balance different classifications by adding weight factors. It assumes that the weight factor added is $\alpha$, and let $\alpha \in [0,1]$, and then it further adds $\alpha$ to the $y = 1$ equation, $1-\alpha$ to the $y = -1$ equation, and it is called $\alpha_t$. The improved balanced cross entropy loss function can be represented as denoted in Eq. (11).

$$L_{bce}(p, y) = -\alpha_t \log(P_t) \tag{11}$$

The balanced cross entropy loss function performs well in addressing the problem of sample imbalance, but it lacks consideration for the difficulty of the sample. Therefore, based on the Focal Loss approach, a modulation factor is introduced into the loss function to solve this problem by aggregating samples that are difficult to distinguish. A weight factor is added, as shown in Eq. (12).

$$\theta = (1-P_t)^{\gamma} \tag{12}$$

Then, the weight factor is added to the loss function for improvement, as shown in Eq. (13).

$$FL(P_t) = -\alpha_t (1-P_t)^{\gamma} \log(P_t) \tag{13}$$

In Eq. (13), $\gamma \in [0,\ 5]$. Furthermore, to maximize the detection ability of the model for WSTs, the idea of size segmentation is studied to further distinguish the WSTs to be detected, and they are divided into large, medium, and small based on their size. Weights are added to WST images, as shown in Eq. (14).

$$\begin{cases} s \in [0,1] \\ S_t = (1-s) \end{cases} \tag{14}$$

Then, according to Eq. (14), it further optimizes the loss function, as shown in Eq. (15).

$$Loss = -\alpha_t (1-P_t)^{\gamma} \log(P_t) \cdot S_t \tag{15}$$

In Eq. (15), $s$ represents the rate of the detected WST image to the overall area of the image. Due to the fact that there is usually no single target to be detected in an image, in the case of non individual targets in the image, this ratio will only calculate the ratio between the smallest target to be detected in the image and the overall image. Through this method, the smaller the image in the image, the greater its ratio, thus strengthening the weight of the weak target in the model training process and thereby enhancing its ability to detect weak targets. The overall structure of the improved SSD-FF weak image detection model constructed by the research institute is shown in Fig. 6.

In Fig. 6, the SSD-based network architecture is studied and Conv3_3 is selected as the new prediction layer with a size of 75 × 75 × 256. The feature pyramid structure is constructed with Conv3_3, Conv4_3 and Conv7 feature layers, and the feature extraction modules are connected with Conv3_3 and Conv4_3 respectively. FF is realized through the feature pyramid structure. After convolution, normalization and ReLU activation, it enters the attention module, and then classification and regression are carried out in the detection layer.

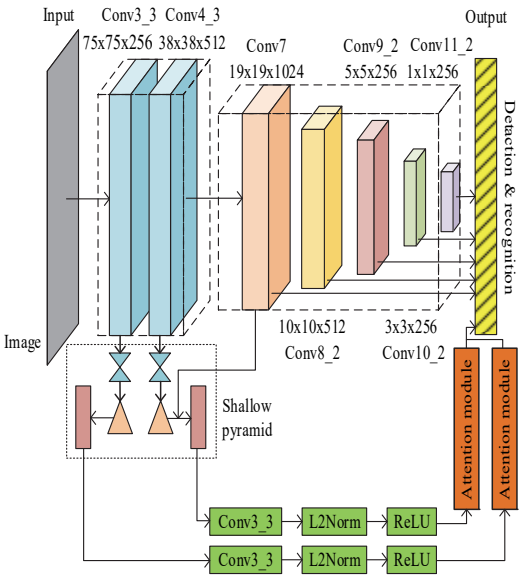


**Figure 6** Structure diagram of SSD-FF WST image detection model

## 3 PERFORMANCE TESTING OF WST IMAGE DETECTION METHOD BASED ON IMPROVED SSD AND DATA FUSION

To validate the effectiveness of the proposed weak and weak image detection model based on improved SSD-FF proposed by the research institute, the PASCAL VOC dataset was selected for experiments. The PASCAL VOC dataset is a widely used standard dataset in computer vision, especially in object detection and image segmentation tasks. It has rich annotation information and real-life image samples, while also possessing high labelling quality. The PASCAL VOC dataset has a large number of categories, which is particularly important for detecting WSTs. The differences in target size and appearance between different categories can promote the generalization ability and make it better adapted to various weak targets. The specific details of the dataset are expressed in Tab. 1. To facilitate subsequent experiments and reduce computational complexity, the study chose to merge the dataset and perform classification processing. All experiments in the study randomly chose 80% of the dataset as the training set and the remaining 20% as the test set.

**Table 1** Data set detail

| Data set | Number of photos | Number of targets | Type of target | The proportion of targets of different sizes | | |
|---|---|---|---|---|---|---|
| | | | | Small size target | Medium size target | Large size target |
| PASCAL VOC 2007 | 9963 | 24640 | 20 | 11.18% | 35.14% | 53.68% |
| PASCAL VOC 2012 | 23080 | 54900 | | | | |

To ensure that the research experiment is not limited by hardware, the model training and experiments in the study were conducted on the server platform. As shown in Tab. 2, the software and hardware details in the study and some parameter settings in the training were presented.

**Table 2** Software and hardware parameters and training parameters

| Name | Detail |
|---|---|
| Experimental framework | Pytorch |
| Development language | Python3.5.2 |
| GPU | Nvidia GeforceRTX 2070 |
| CPU | Intel Core(R) i7-7700k × 4.2 Ghz |
| RAM | 8 G × 2666 Mhz × 2 |
| Epoch | 50 |
| Batch size | 8 |
| Learning rate | 10-4 |
| IoU threshold | 0.5 |

In the study, improvements were made to the basic SSD network to maximize its ability to detect WST images, with three main points. The first point was the addition of specialized prediction heads for WSTs. The second point was the improved shallow feature pyramid structure. The third point was the context attention FF module based on FF. To validate the performance of the improvements made in the study, experiments were organized to compare the effects of different modules. The baseline model in the experiment was the basic SSD network, where Model 1was an SSD network with only a prediction head added; Model 2 was an SSD network with added prediction heads and improved shallow feature pyramid structure; Model 3 was an SSD network with a prediction head and a context attention FF module added; Model 4 was an SSD network with added prediction heads, improved shallow pyramid structure, and contextual attention FF modules. The research findings are denoted in Tab. 3. From Tab. 3, the proposed improved method has improved the detection effectiveness of large, small, and medium-sized targets in SSD networks, but only the combined method had the best effect, with a small target detection accuracy of 34.69%, which was 26.54% higher than the SSD model. The average accuracy reached 82.39%, which was 11.05% higher than the SSD model.

**Table 3** Results of ablation experiments

| Model | Target detection accuracy rate | | | |
|---|---|---|---|---|
| | Small size target detection accuracy rate | Medium size target detection accuracy rate | Large size target detection accuracy rate | Average target detection accuracy |
| Baseline | 8.15% | 41.27% | 72.15% | 71.34% |
| Model 1 | 17.92% | 39.81% | 75.26% | 72.15% |
| Model 2 | 21.38% | 42.39% | 76.64% | 73.38% |
| Model 3 | 25.34% | 46.58% | 75.38% | 73.47% |
| Model 4 | 34.69% | 51.27% | 83.68% | 82.39% |

The research further selected dim target detection algorithm to compare the detection accuracy with the method proposed in the research, specifically the Corner-Net algorithm and Faster R-CNN algorithm, in which the Corner-Net algorithm is a one-stage target detection algorithm [29] and Faster R-CNN is a two-stage target detection algorithm [30]. After training all three algorithms to their optimal state, they were tested for accuracy using an integrated dataset and further tested for their response time performance. The test results are shown in Fig. 7. From Fig. 7, the proposed SSD-FF algorithm had higher recognition accuracy in all categories than the other two algorithms, with an average of 47.67% and 79.56% higher than the Corner-Net and Faster R-CNN algorithms, respectively. In response time, the optimal response time of the proposed algorithm was 0.62 seconds, which was

0.91 seconds ahead of the Corner-Net algorithm and 0.97 seconds ahead of the Faster R-CNN algorithm, respectively.

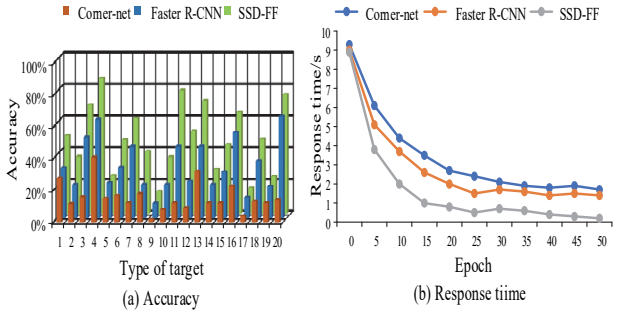


**Figure 7** Detection accuracy and response time test results

The proposed SSD-FF was compared with the baseline network SSD for target recognition, and the test outcomes are expressed in Fig. 8. From Fig. 8, both methods had good sensitivity to large targets, but SSD could not recognize smaller targets in the distance. In comparison, the SSD-FF algorithm could effectively detect small size targets in the distance, and its detection ability for conventional size targets in close range was also excellent [31].


(a) SSD algorithm detection results (b) SSD-FF algorithm detection results

**Figure 8** Comparison chart of test results

Due to the focus of the research on visual communication, a comparison was made between the visual communication processing time of the proposed algorithm and the visual communication technology based on the Corner-Net and the Faster R-CNN algorithms [32]. To fully validate the visual communication performance of the proposed method, the dataset was filtered based on image pixels, and the visual communication processing time of the algorithm was tested and compared with changes in image resolution. The comparison findings are displayed in Fig. 9. From Fig. 9, the proposed method had a shorter processing time, leading by 9 seconds and 19 seconds compared to the Corner-Net algorithm and Faster R-CNN algorithm, respectively.

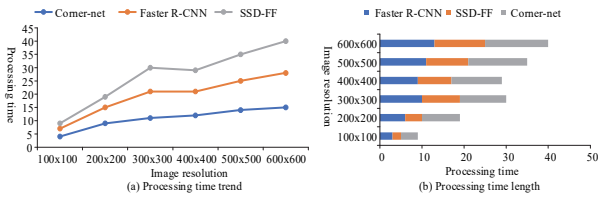


**Figure 9** Visually communicate the results of processing time comparison

Finally, the visual communication success rate of the raised method was tested, and the test outcomes are shown in Fig.10. From Fig. 10, the visual communication success rate of all methods fluctuated with the number of image frames. However, the proposed method had the highest visual communication success rate, with an average of 98%, which was 16% and 27% higher than the Corner-Net algorithm and Faster R-CNN algorithm, respectively. In summary, the SSD-FF WST image detection algorithm had extremely high accuracy in identifying WSTs, and in actual visual communication, it had shorter visual communication processing time and higher success rate in visual communication [33].

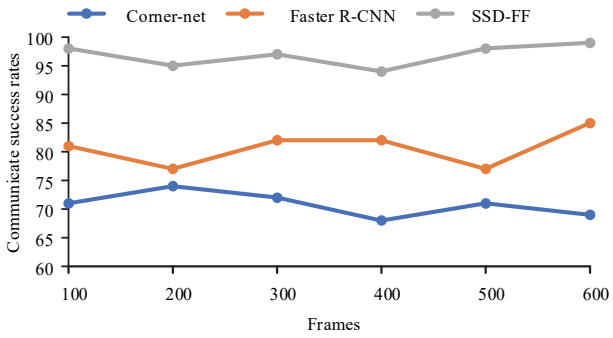


**Figure 10** Visually communicate the results of success

## 4 CONCLUSION

The field of computer vision is of great significance in modern technology and society. It is a discipline that studies and develops the use of computers to simulate and improve human visual perception and understanding abilities. To overcome the difficulty of detecting WST images in the field of computer vision, this study presented a visual communication design using an improved SSD architecture and FF to enhance WST detection. The proposed approach achieved significant gains in detection accuracy, faster response times, and higher visual communication success versus state-of-the-art methods. The specialized model components helped overcome major challenges with small, low contrast targets. Detailed experiments demonstrated the feasibility and potential of the design. The results of the ablation experiment proved the effectiveness of the improvements made by the research institute. In contrast with the original model, the improved model achieved a detection accuracy of 34.69% for WSTs, with an improvement of up to 26.54%. The average accuracy reached 82.39%, with an improvement of 11.05%. Further experiments showed that the algorithm proposed by the research institute had higher recognition accuracy on all categories of targets compared to current advanced algorithms, with an average of 47.67% and 79.56% higher than the Corner-Net algorithm and Faster R-CNN algorithm. This aligned with [20] and highlighted the value of fusion for handling WST challenges. The proposed algorithm also had a shorter response time, with the best response time reaching 0.62 s, and its performance was consistent with [18], which highlighted the value of its improved detection speed for WST. In experiments on visual communication, the algorithm proposed by the research institute had better visual communication time, which was 9 seconds and 19 seconds higher than the Corner-Net algorithm and Faster R-CNN algorithm, respectively. The proposed method also had a better visual communication success rate, with an average success rate of 98%, which was 16% -27% ahead of the two algorithms. In summary, compared with traditional dim and small

image detection algorithms, the proposed dim and small image detection model based on improved SSDFF has excellent performance in the field of weak and small image detection, with high weak and small target recognition accuracy, short response time, and higher visual transmission efficiency and success rate. Overall, this work makes a key contribution to weak and small target detection and visual communication of key weak and small image targets. However, only one dataset was used for training and research in the study, lacking further sample diversity, which should be improved in subsequent studies. Future work should focus on extending the approach to more complex datasets and targets.

## 5 REFERENCES


[1] Ratnayake, M. N., Amarathunga, D. C., Zaman, A., Dyer, A. G., & Dorin, A. (2023). Spatial Monitoring and Insect Behavioural Analysis Using Computer Vision for Precision Pollination. *Journal of Computer Vision*, *131*(5), 1300-1301. https://doi.org/10.1007/s11263-022-01715-4

[2] Zhang, Y. & Lin, W. (2021). Computer-vision-based differential remeshing for updating the geometry of finite element model. *Computer-Aided Civil and Infrastructure Engineering*, *37*(2), 185-203. https://doi.org/10.1111/mice.12708

[3] Xie, J., Huang, S., Wei, D., & Zhang, Z. (2022). An infrared dim small target detection algorithm based on spatial and temporal fusion of mathematical morphology. *International Conference on High Performance Computing and Communication (HPCCE 2021)*, 190-195.

[4] Macdonald, I. (2023). Window on the weather: a case study in multi-platform visual communication design, with a relationship to Design Thinking. *Visual Communication*, *22*(2), 365-386. https://doi.org/10.1177/1470357220948547

[5] Plant, S., Lundin, K., & Alvesson, H. M. (2022). 'It touches my heart more when I see this': visual communication in the realization of risk - the case of type 2 diabetes in Stockholm. *Health, Risk & Society*, *24*(5/8), 258-275. https://doi.org/10.1080/13698575.2022.2104221

[6] Wang, J. (2021). Application of wavelet transform image processing technology in financial stock analysis. *Journal of Intelligent and Fuzzy Systems*, *40*(2), 2017-2027. https://doi.org/10.3233/JIFS-189204

[7] Navi, K., Uoosefian, H., Mirzaee, R. F., & Hosseinzadeh, M. (2020). High-Performance CML approximate full adders for image processing application of Laplace transform. *Circuit World*, *46*(4), 285-299. https://doi.org/10.1108/CW-12-2018-0106

[8] Xia, D. H., Song, S. Z., Tao, L., Qin, Z. B., Wu, Z., Gao Z. M., Wang J. H., Hu W. B., Behnamian, Y., & Luo, J. L. (2020). Review-material degradation assessed by digital image processing: Fundamentals, progresses, and challenges. *Journal of Materials Science & Technology*, *53*(18), 148-164. https://doi.org/10.1016/j.jmst.2020.04.033

[9] Khaddam, H. S. & Ahmad, G. G. (2022). A method to evaluate the diameter of carded cotton yarn using image processing and artificial neural networks. *The Journal of the Textile Institute*, *113*(8), 1648-1657. https://doi.org/10.1080/00405000.2021.1943259

[10] Lopez, N. R., Tao, Y., Elik, H., & Hopmann, C. (2023). Development of an image processing algorithm (IPA - Delfin) for the digital reconstruction of composite overwrapped pressure vessels. *Polymer Composites*, *44*(4), 2417-2426. https://doi.org/10.1002/pc.27253

[11] Yang, B. & Wang, J. (2022). An Improved Helmet Detection Algorithm Based on YOLO V4. *International Journal of Foundations of Computer Science*, *33*(6/7), 887-902. https://doi.org/10.1142/S0129054122420205

[12] Zhao, M., Yue, L., Hu, J., Du, S., Li, P., & Wang, L. (2021). Salient target detection in hyperspectral image based on visual attention. *IET Image Processing*, *15*(10), 2301-2308. https://doi.org/10.1049/ipr2.12197

[13] Sun, M., Zhang, H., Huang, Z., Luo, Y., & Li, Y. (2022). Road infrared target detection with I-YOLO. *IET Image Processing*, *16*(1), 92-101. https://doi.org/10.1049/ipr2.12331

[14] Zhang, T., Peng, Z., Wu, H., He, Y., Li, C., & Yang, C. (2021). Infrared small target detection via self-regularized weighted sparse model. *Neurocomputing*, *420*, 124-148. https://doi.org/10.1016/j.neucom.2020.08.065

[15] Zhang, L., Li, M., Qiu, X., & Zhu, Y. (2020). Infrared Small Target Detection Based on Four-Direction Overlapping Group Sparse Total Variation. *Traitement du Signal*, *37*(3), 367-377. https://doi.org/10.18280/ts.370303

[16] Xing, Z., Chen, X., & Pang, F. (2022). DD-YOLO: An object detection method combining knowledge distillation and Differentiable Architecture Search. *IET Computer Vision*, *16*(5), 418-430. https://doi.org/10.1049/cvi2.12097

[17] Yan, J. & Wang, Z. (2022). YOLO V3+VGG16-based automatic operations monitoring and analysis in a manufacturing workshop under Industry 4.0. *Journal of Manufacturing Systems*, *63*(1), 134-142. https://doi.org/10.1016/j.jmsy.2022.02.009

[18] Guo, L., Wang, D., Li, L., & Feng, J. (2020). Accurate and fast single shot multibox detector. *IET Computer Vision*, *14*(6), 391-398. https://doi.org/10.1049/iet-cvi.2019.0711

[19] Wang, G., Ding, H., Li, B., Nie, R., & Zhao, Y. (2022). Trident-YOLO: Improving the precision and speed of mobile device object detection. *IET image processing*, *16*(1), 145-157. https://doi.org/10.1049/ipr2.12340

[20] Yang, S., Chen, Z., Ma, X., Zong, X., & Feng, Z. (2022). Real-time high-precision pedestrian tracking: a detection-tracking-correction strategy based on improved SSD and Cascade R-CNN. *Journal of Real-Time Image Processing*, *19*(2), 287-302. https://doi.org/10.1007/s11554-021-01183-y

[21] Barbhuiya, A. A., Karsh, R. K., & Jain, R. (2022). A convolutional neural network and classical moments-based feature fusion model for gesture recognition. *Multimedia systems*, *28*(5), 1779-1792. https://doi.org/10.1007/s00530-022-00951-5

[22] Song, Y. & Fan, C. (2023). Behavior Recognition of the Elderly in Indoor Environment Based on Feature Fusion of Wi-Fi Perception and Videos. *Journal of Beijing Institute of Technology*, *32*(2), 142-155. https://doi.org/10.15918/j.jbit1004-0579.2022.131

[23] Zhang, R., Zhao, N., Fu, L., Pan, L., Bai, X., & Song, R. (2022). Ultrasonic diagnosis method for stainless steel weld defects based on multi-domain feature fusion. *Sensor Review*, *42*(2), 214-229. https://doi.org/10.1108/SR-08-2021-0272

[24] Cui, X., Xu, Y., Lou, Y., Sheng, Q., Cai, M., & Zhuang, L. (2022). Diagnosis of Parkinson's disease based on feature fusion on T2 MRI images. *International journal of intelligent systems*. https://doi.org/10.1002/int.23046

[25] Wang, L., Li, Y., Peng, S., Tang, X., & Yin, B. (2021). Multi-level feature fusion network for crowd counting. *IET Computer Vision*, *15*(1), 60-72. https://doi.org/10.1049/cvi2.12012

[26] Zhao, M., Zhong, Y., Sun, D., & Chen, Y. (2021). Accurate and efficient vehicle detection framework based on SSD algorithm. *IET Image Processing*, *15*(13), 3094-3104. https://doi.org/10.1049/ipr2.12297

[27] Lee, Y. S. (2022). A study on abnormal behavior detection in CCTV images through the supervised learning model of deep learning. *Journal of Logistics, Informatics and Service Science*, *9*(2), 196-209. https://doi.org/10.33168/LISS.2022.0212

[28] Barbhuiya, A. A., Karsh, R. K., & Jain, R. (2022). A convolutional neural network and classical moments-based

feature fusion model for gesture recognition. *Multimedia Systems*, *28*(5), 1779-1792. https://doi.org/10.1007/s00530-022-00951-5

[29] Gomaa, M. M., Mohamed, E. R., Zaki, A. M., & Elnashar, A. (2022). Deep learning to detect image forgery based on image classification. *Journal of System and Management Sciences*, *12*(6), 454-467. https://doi.org/10.33168/JSMS.2022.0628

[30] Oslund, S., Washington, C., So, A., Chen, T., & Ji, H. (2022). Multiview Robust Adversarial Stickers for Arbitrary Objects in the Physical World. *Journal of Computational and Cognitive Engineering*, *1*(4), 152-158. https://doi.org/10.47852/bonviewJCCE2202322

[31] Zhao, G., Dong, T., & Jiang, Y. (2022). Corner-based object detection method for reactivating box constraints. *IET Image Processing*, *16*(13), 3446-3457. https://doi.org/10.1049/ipr2.12576

[32] Zhang, D., Zhan, J., Tan, L., Gao, Y., & Župan, R. (2021). Comparison of two deep learning methods for ship target recognition with optical remotely sensed data. *Neural Computing and Applications*, *33*(10), 4639-4649. https://doi.org/10.1007/s00521-020-05307-6

[33] Asriny, D. M. & Jayadi, R. (2023). Transfer learning VGG16 for classification of orange fruit images. *Journal of System and Management Sciences*, *13*(1), 206-217. https://doi.org/10.33168/JSMS.2023.0112

**Contact information:**

**Xu ZHANG**
(Corresponding author)
Wuhan University of Communication,
Wuhan, Hubei, China, 430205
E-mail: xiaoxuer25@163.com