electronics

MDPI

Article

FEFD-YOLOV5: A Helmet Detection Algorithm Combined with Feature Enhancement and Feature Denoising

Yiduo Zhang †, Yi Qiu † and Huihui Bai *

Institute of Information Science, Beijing Jiaotong University, Beijing 100044, China; 21140124@bjtu.edu.cn (Y.Z.); 21140076@bjtu.edu.cn (Y.Q.)

* Correspondence: hhbai@bjtu.edu.cn

† These authors contributed equally to this work.

Abstract: In intelligent surveillance of construction sites, safety helmet detection is of great significance. However, due to the small size of safety helmets and the presence of high levels of noise in construction scenarios, existing detection methods often encounter issues related to insufficient accuracy and robustness. To address this challenge, this paper introduces a new safety helmet detection algorithm, FEFD-YOLOV5. The FEFD-YOLOV5 algorithm enhances detection performance by adding a shallow detection head specifically for small target detection and incorporating an SENet channel attention module to compress global spatial information, thus improving the model's mean average precision (mAP) in corresponding scenarios. Additionally, a novel denoise module is introduced in this algorithm, ensuring the model maintains high accuracy and robustness under various noise conditions, thereby enhancing the model's generalization capability to meet real-world scenario demands. Experimental results show that the proposed improved algorithm achieves a detection accuracy of 94.89% in noiseless environments, and still reaches 91.55% in high-noise environments, demonstrating superior detection efficacy compared to the original algorithm.

Keywords: safety helmet; small target; attention module; YOLOv5s; denoise module

check for updates

Citation: Zhang, Y.; Qiu, Y.; Bai, H. FEFD-YOLOV5: A Helmet Detection Algorithm Combined with Feature Enhancement and Feature Denoising. *Electronics* **2023**, *12*, 2902. https://doi.org/10.3390/electronics12132902

Academic Editor: Hüseyin Kusetogullari

Received: 5 June 2023
Revised: 26 June 2023
Accepted: 26 June 2023
Published: 1 July 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

As urban construction rapidly advances, the safety concerns at construction sites have gained increasing attention. Safety helmets, as part of the protective gear on construction sites, play a crucial role in ensuring worker safety. However, due to the small size of safety helmets and the substantial amount of noise present in construction scenes, traditional helmet detection methods demonstrate inadequacies in terms of accuracy and robustness. In recent years, significant advancements in deep learning technologies within the field of object detection have offered new possibilities for improving safety helmet detection algorithms.

The field of object detection has made significant strides, particularly where methods based on deep learning have achieved breakthrough improvements in performance across numerous tasks. Presently, mainstream object detection methods primarily fall into two categories: region-based methods (such as Faster R-CNN) and single-stage methods (such as SSD and YOLO series). Faster R-CNN employs a region proposal network (RPN) to generate candidate regions, utilizes convolutional neural networks (CNN) for feature extraction, and eventually applies a fully connected layer for feature classification and bounding box regression. However, Faster R-CNN's computational complexity escalates when dealing with a multitude of candidate regions, leading to slower detection speeds. SSD achieves higher detection speeds by predicting on multi-scale feature maps, but its performance is somewhat lacking in small target detection. The YOLO series transforms the object detection problem into a regression problem, achieving the goal of real-time detection, albeit with slightly reduced detection accuracy compared to Faster R-CNN.

Given the specific task of safety helmet detection, the YOLO algorithm has advantages in terms of faster detection speed and easier model deployment. However, it exhibits certain deficiencies in terms of accuracy in complex and open scenarios. In real-world scenes, detection accuracy significantly varies for objects of different sizes. Moreover, construction site scenarios present noise issues due to factors like fog, wind-blown dust, dynamic blurring and occlusion from construction activities, and even artifacts from the camera itself, affecting the model's recognition rate and detection speed. To address issues related to small target detection and robustness under noise conditions, this paper introduces a safety helmet detection algorithm based on feature enhancement and noise reduction (FEFD-YOLOV5). Experimental results show that the proposed FEFD-YOLOV5 algorithm outperforms YOLOv5s in terms of mean average precision (mAP) under both noiseless and noisy conditions, thus validating its effectiveness.

The main contributions of this paper are summarized as follows:

1. A shallow object detection head is added to the prediction part, integrating shallow features comprehensively. This method effectively leverages the shallow features of small objects, enhances the semantic representation of small objects, and improves the detection performance of the safety helmet detection model.
2. By introducing an attention module, the algorithm selectively learns and adjusts the interdependencies among different channels in the feature map. This approach highlights the relevant information of small objects and enhances their representation in the feature space, which in turn, significantly improves the detection accuracy of small objects in complex scenarios.
3. A novel noise reduction block is proposed based on the Non-Local Means algorithm, eliminating noise within features extracted by the convolutional layer, thereby improving detection results.

2. Related Works

The application of deep learning in the field of helmet detection has gradually become a research focus. Computer vision algorithms can be primarily divided into single-stage and two-stage detection types. In single-stage detection, the representative ones include the YOLO series, from YOLOv1 to YOLOv5 [1–4], the SSD series, including the rainbow SSD [5] (R-SSD), feature-fusion SSD [6] (FSSD), and deconvolutional SSD [7] (DSSD), as well as EfficientDet [8] among others. In two-stage detection, region-based convolutional neural networks (R-CNNs), such as R-CNN [9], Fast R-CNN [10], Faster R-CNN [11], and Mask r-cnn [12], hold dominant positions.

However, both single-stage and two-stage detection algorithms may suffer from the loss of shallow features due to the depth of network training, which can gradually weaken the semantic information of small targets. To address this issue, researchers have attempted to combine shallow features with deep features. Approaches such as Feature Pyramid Network [13] (FPN), Path Aggregation Network [14] (PAN), Parallel Residual Bi-directional Fusion Feature Pyramid Network [15] (PRB-FPN), and Receptive Field Block [16] (RFB) are representative attempts. These methods ensure the integrity of feature information by extracting features from shallow layers, which can somewhat alleviate the issue of small target detection.

In addition, to further enhance model performance, researchers have tried to integrate denoising modules, such as Non-local Neural Networks [17]. This algorithm effectively reduces noise interference while processing global contextual information. Some research has also focused on using attention mechanisms to enhance model performance, like Convolutional Block Attention Module [18] (CBAM), and Squeeze-and-Excitation Networks [19] (SENet), which help the model focus more on essential information.

For instance, Kun Han et al. [20] currently used multi-scale detection to identify helmets, effectively improving recognition accuracy by adding a fourth dimension to predict a large number of small targets. Li N et al. [21] proposed a helmet detection method based on a modified shallow feature network for Faster RCNN and a multi-part fusion.

Likewise, Fan Wu et al. [22] effectively handled cases where helmets were stained, partially obscured, or where there were numerous targets in low-resolution images by switching the backbone feature extraction network. Moreover, G Han et al. [23] applied channel attention modules to high-level features to better capture global contextual information, facilitating the extraction of crucial feature channels. By introducing feature pyramids and multi-scale perception modules, the robustness of the detection algorithm to target scale changes was improved. Weipeng Tai et al. [24] enhanced the features of the spatial region corresponding to the target object by introducing a spatial attention module, thereby improving helmet recognition rate.

Base on the aforementioned research, we have extensively optimized the architecture of YOLOv5. Our aim is to significantly improve detection accuracy and efficiency, particularly with regard to noise interference and small targets. This section outlines the core contributions of our work, highlighting our unique innovations and improvements in the field of helmet detection.

3. FEFD-YOLOv5

3.1. Network Structure

The network structure of FEFD-YOLOV5, shown in Figure 1, mainly consists of the Backbone, Neck, and Head. Compared to YOLOv5s, FEFD-YOLOV5 has made three significant improvements to enhance feature extraction capability and mitigate the influence of noise:

1. Addition of a small object detection head to the shallow feature section: This is specifically designed for small objects, with an input size of $160 * 160 * 255$. Its primary function is to capture the information of small objects in the early stages of feature extraction, ensuring that this information is not lost or blurred in the subsequent deeper network layers.
2. Incorporation of the SENet attention module, forming an S3 structure: By introducing an attention mechanism into the network, the importance of feature channels can be dynamically adjusted. The SENet module can automatically learn and extract features that are more useful for object recognition, thereby concentrating the model's focus and reducing the interference of irrelevant information.
3. Addition of a noise reduction module to the Backbone: The Backbone, being the main part of the network, is used to obtain the basic features of the original image. Adding a noise reduction module helps eliminate image noise at an early stage, enhancing the model's resistance to noise and ensuring the clarity and accuracy of the features.

The original image, with a size of $640 * 640 * 3$, first passes through the noise reduction module for preprocessing, eliminating the original image's noise and extracting preliminary feature information. These features pass to the S3 module, which further extracts and filters out features that are more useful for object detection using the SENet attention mechanism. Simultaneously, the small object detection head in the shallow feature section preserves and extracts fine-grained information that could potentially be lost in subsequent deeper network layers. Through these three improvements, FEFD-YOLOV5 effectively recognizes small objects and resists noise, significantly enhancing the accuracy of safety helmet detection.

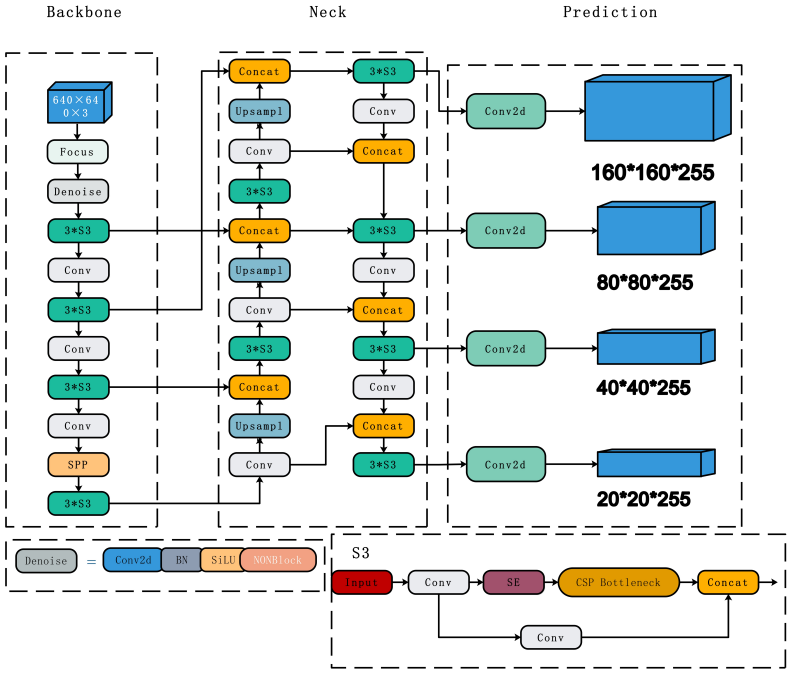


Figure 1. FEFD-YOLOV5 Model Structure.

3.2. Small Object Detection Layer

The detection of small objects is a significant challenge in deep learning object detection algorithms. Because the increase in network depth and the downsampling process of the feature map often leads to the loss of precise information about the position of small objects, blurred location information directly affects the recognition accuracy of small objects. In this context, we have researched and designed a strategy specifically optimized for small object detection.

Specifically, we have added a new detection head to the existing YOLOv5s model, which is specifically used for the recognition of small objects. This new small object detection layer processes images at a higher resolution in the shallow network layers, effectively capturing detailed information about small objects (such as safety helmets). We leverage an important phenomenon where small objects in high-resolution images can produce more feature points, thereby significantly enhancing the model's recognition performance. Since this detection head only handles small objects, the original deep detection head is still used for processing large objects. Therefore, the new network structure maintains computational efficiency while effectively improving the recognition ability for small objects.

The network structure presented in Figure 2 incorporates newly added layers specifically for small object recognition, denoted by red lines and red boxes. We introduced a new small object recognition layer into the shallow structure of the original YOLOv5 model, while retaining the mutual fusion structure strategy of the Feature Pyramid Network (FPN) and the Path Aggregation Network (PAN). In the Backbone part, the feature map first passes through the feature extraction network module, then enters the Neck part, and fuses with the upsampled feature map. After another feature extraction network module, the feature map enters the detection layer. Meanwhile, part of the feature map is downsampled and fused with the feature map from another path, thereby effectively enhancing the location information of shallow features.

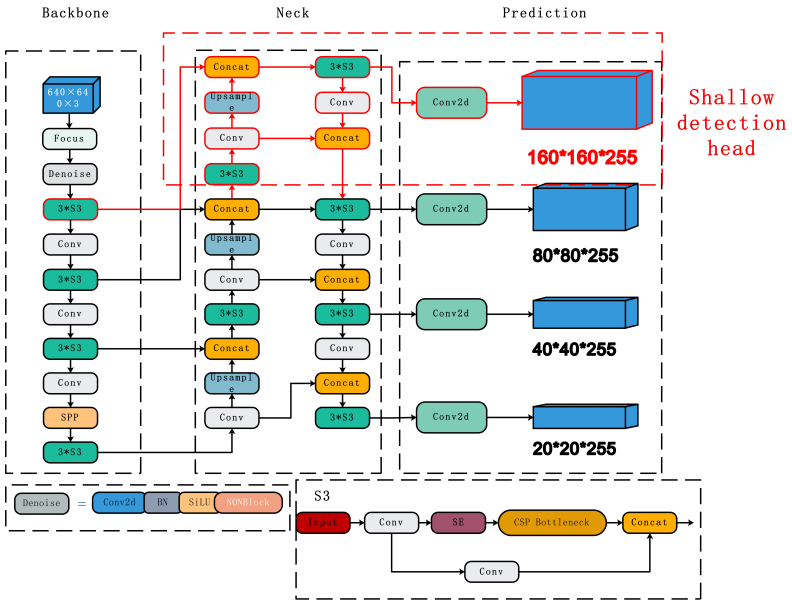


Figure 2. The shallow object detection architecture. The red lines and boxes denote components that are added compared to the original version.

By adopting this improved strategy, we can fully utilize the location information in the shallow feature map, significantly enhancing the model's performance in detecting small-sized safety helmets. Through experimental verification, we have found that this improvement strategy indeed shows significant advantages in handling small-sized safety helmet detection tasks, contributing to the accuracy and robustness of the model in real-world application scenarios.

3.3. Channel Attention Module

To further enhance the performance of our model in the safety helmet detection task, we incorporated a channel attention module into the model's C3 module. This module, based on the concept of SENet, intensifies the focus on crucial features by mining inter-channel relationships in the feature maps, thereby boosting the model's recognition capability.

As a classic channel attention model, SENet has significant application value in the field of computer vision. SENet mainly includes two parts: Squeeze and Excitation. The Squeeze part is responsible for compressing global spatial information to facilitate feature learning at the channel dimension, while the Excitation part employs a method similar to a multi-layer perceptron [25] (MLP) to model channel importance, achieving channel weighting and normalization. After multiplying the feature maps by the channel weights, the importance of each channel in the feature maps can be quantified, further enhancing model performance.

Figure 3 displays the structure of the SE module. A standard 2D convolution operation is initially used to transform the input feature map into feature maps of varying sizes. Global average pooling and global maximum pooling are two common methods of Squeeze operations. This study utilizes global average pooling as the pooling method. Through calculation Equation (1), a one-dimensional vector is obtained:

$$Avg_i(x) = \frac{1}{H_i \times W_i} \sum_{k=1}^{H_i \times W_i} f_{ik}(i = 1 \cdots c) \quad (1)$$

The Excitation operation is a fully connected layer operation aiming to reduce the parameter scale of the attention network, thus realizing model lightweightness and efficiency. After processing by a Sigmoid activation function, we obtain weights for C channels,

distributed between 0 and 1. Then, these weights are combined with the transformed feature maps, enabling the model to distinguish the importance between channels. The larger the weight, the more significant the impact of the corresponding feature channel on the final target recognition.

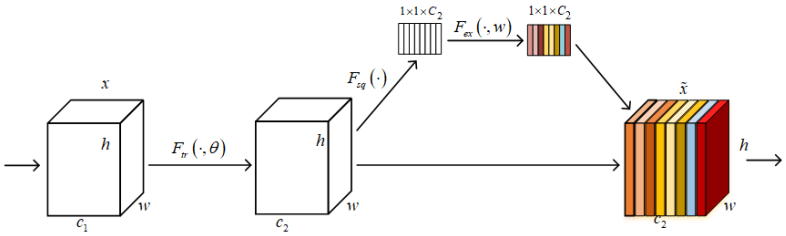


Figure 3. The architecture of SE network.

To enhance the model's performance in the extraction of small target features, we incorporated SENet into the feature extraction stage, enabling the model to better distinguish between targets and the background, thereby improving small target recognition ability.

Figure 4 demonstrates the network structure with the attention module. The input feature map is 64 × 160 × 160. After the attention module assigns different channel or spatial weights to the features, we obtain a feature map of the same size, 64 × 160 × 160. Then, we downsample the feature map with attention weights, and input the feature map obtained from the previous layer through convolution into the residual branch, thereby extracting deeper feature information. The design of the Bottleneck layer structure aims to reduce the number of parameters, decrease computation, and enhance data training and feature extraction efficiency after dimension reduction. Finally, we perform a concat operation on the two feature maps to obtain the final output.

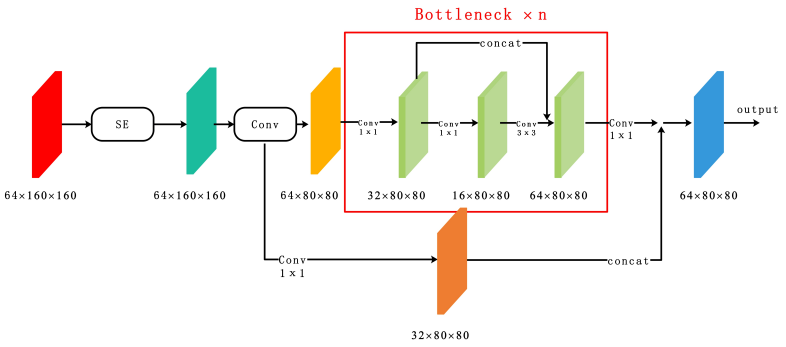


Figure 4. Network Structure with the Attention Module.

Upon incorporating the channel attention module, the model's performance in the safety helmet detection task was significantly enhanced. The model's accuracy improved significantly, and its small target recognition capability was notably enhanced. Based on this, our work confirmed that introducing the channel attention module into the target detection model can effectively improve the model's performance, enabling it to have higher accuracy and robustness in real-world application scenarios.

3.4. Denoise Module

When constructing a safety helmet detection model, besides improving model recognition accuracy through enhanced channel learning, enhancing the non-local correlation in images is also an important task. Detection failures sometimes stem from the model's over-interpretation of background noise in images. Therefore, we proposed a feature denoising method that combines the Non-Local Means (NLM) algorithm and convolutional neural networks. This method takes full advantage of the correlations between image information, effectively suppresses image noise, and improves model performance.

Unlike traditional denoising in the image preprocessing stage, our method goes further by denoising feature maps in the convolutional neural network. The designed noise reduction module uses a Gaussian filter, fully considering the spatial information at each location and utilizing the redundant information and structural information in the image to eliminate noise. This allows the model to retain image details and texture while reducing noise.

The definition of the denoising block function is shown in Equation (2):

$$y_i = \frac{1}{C(\hat{x})} \sum_{\forall j} w(x_i, \hat{x}_j) v(\hat{x}_j), \tag{2}$$

where i represents the output location, such as spatial, temporal, or spacetime indexes, whose response is obtained by enumerating and calculating j. X is the input feature map, y is the denoised feature map, and $\hat{x}_j$ is the feature obtained by downsampling. The normalization operation $C(\hat{x})$ performs using Softmax. The function $v(\hat{x}_j)$ calculates the representation of the feature map at position j, and is defined in Equation (3):

$$v(\hat{x}_j) = W\hat{x}_j, \tag{3}$$

in which, the matrix W is obtained by training and updating the convolution of 1×1.

The denoising function realized by the Gaussian filter $W(x)$ is defined as shown in Equation (4):

$$W(x_i, \hat{x}_j) = e^{x_i^T x_j}. \tag{4}$$

Here, a dot product $x_i^T x_j$ is used to calculate the similarity, which is simplified from the cosine similarity. Compared to the commonly used Euclidean distance calculation method, using the dot product operation has more advantages in a neural network environment. When dealing with images disturbed by noise, this Gaussian function can effectively reduce the influence of noise, thereby enhancing image quality.

As shown in Figure 5, the construction of the denoising module, in the case where the size of the input feature map is $H \times W \times 64$, uses a 1×1 convolution for channel transformation and matrix transformation, and carries out the denoising operation through matrix multiplication. Specifically, at the three positions of a, b, and c, a 1×1 convolution is used for channel transformation, reducing the size of the feature map by half, aiming to reduce the computational load of the module and improve the efficiency of the module. To achieve the purpose of denoising, we use the Softmax function for normalization, and perform a weighted average operation on the entire image position. The denoised feature map is added to the original feature map through residual connection, and then a 1×1 convolution is used to obtain a feature map consistent with the initial shape.

The design of this residual structure allows for feature fusion without loss of input information, enabling the network to learn various complex features and representations more effectively. This design strategy not only preserves more beneficial information, but also equips the network with the ability to handle more complex tasks, thereby significantly enhancing network performance.

In summary, the feature denoising method we proposed, which combines the Non-Local Means (NLM) algorithm and convolutional neural networks, not only effectively eliminates image noise and improves image quality, but also significantly reduces the false alarm rate in complex scenarios, thereby achieving the goal of improving model performance while maintaining high efficiency. This is achieved by employing a Gaussian filter in the denoising module and effectively integrating complex features.

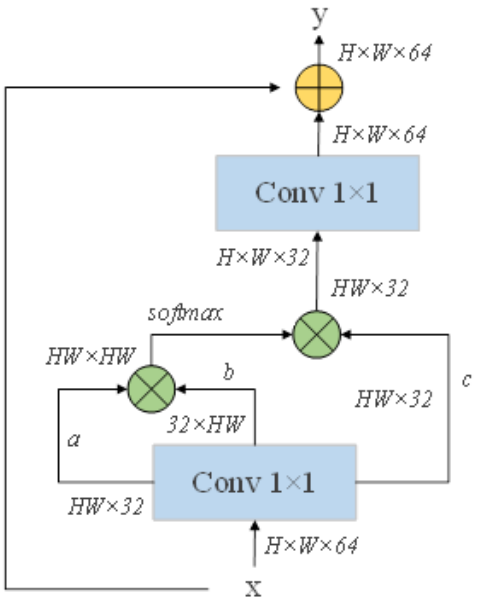


Figure 5. Structure of the Denoising Block.

4. Experimental Results and Analysis

4.1. Experimental Datasets

In this section, we utilized a dataset comprised of 6140 images for evaluating the proposed algorithm. The images in the dataset include a combination of custom data captured by cameras at different construction site scenes during various days, under diverse lighting and weather conditions, and a portion of safety helmet wearing images obtained from the Internet. This Internet-derived dataset includes images from multiple sources to ensure the variability and representativeness of the data.

The collected images were manually annotated with bounding boxes for the helmet objects and tagged with class labels. The annotation process was carried out by three independent individuals to ensure the accuracy of the labels.

We partitioned the annotated dataset into training, validation, and test sets in a ratio of 8:1:1, ensuring that each set contained varied representative samples from both the custom and Internet-derived datasets. The training set is used for learning the model parameters, the validation set for tuning the hyperparameters, and the test set for evaluating the generalization ability of the proposed FEFD-YOLOV5 algorithm.

To evaluate the performance of the algorithm under noise in the experiment, we chose a type of noise that is quite close to the natural environment—Gaussian noise—for noise addition processing of the test images. This type of noise has been widely used in the field of image denoising research. Additive White Gaussian Noise (AWGN) follows a Gaussian distribution. This noise model helps to more accurately evaluate the performance of the proposed algorithm in actual application scenarios. Its mathematical representation is shown in Equation (5):

$$f(x) = \frac{1}{\sqrt{2\pi\sigma}} exp\left(-\frac{x^2}{2\sigma^2}\right) \quad (5)$$

In the equation, σ represents the standard deviation. The strength of the noise can be effectively controlled by adjusting the size of the standard deviation. AWGN is considered an ideal noise signal because it can uniformly add noise values to each pixel value and simulate the noise caused by factors such as wireless transmission and signal interference in the actual environment within specific frequency bands. Due to its controllability and ease of analysis and comparison, image denoising algorithms usually use AWGN to simulate

the noise situation in real environments. Figure 6 shows a comparison of the image before and after adding noise at the $\sigma = 15$ level.



(**a**) original image (**b**) Gaussian white noise

Figure 6. Comparison of the original image and the image after adding Gaussian white noise.

In this experiment, the level of noise added to the training set is $\sigma = 15$. For the test set, it is tested in an environment with a noise level of $\sigma = 0, 15, 25, 35, 50$.

4.2. Experimental Environments

In this study, we set the epoch to 300, batch-size to 16, and the size of the input images to 640×640. The training process spanned the first four epochs, utilizing a stepwise increasing strategy. The learning rate for weight updates started from 0 and increased arithmetically to 0.02, while the bias layer was gradually reduced from 0.1 to 0.02. After completing the training over the initial four epochs, an adaptive learning rate adjustment strategy was utilized to update the learning rate (Table 1).

Table 1. Experimental Configuration.

Hardware	Software
Graphics Card	3090
Graphics Memory	24 GB
RAM	32 GB
Hard Drive	20 TB
Operating System	Ubuntu 18.04.6
Development Environment	CUDA 11.4
Programming Language	Python 3.6
Framework	PyTorch

4.3. Experimental and Evaluation Criteria

At the onset of the learning phase, the system utilizes the training set, which is designed to enable the model to learn the model parameters. The initial four epochs are subjected to a learning process using a stepwise increasing strategy. The learning rate for weight updates starts from 0 and increased arithmetically to 0.02, while the bias layer gradually reduces from 0.1 to 0.02. In essence, this process allows the model to understand the patterns, trends, and correlations within the training data, thereby establishing the underlying model for the algorithm.

Subsequent to the learning phase, the algorithm proceeds to the optimization phase. In this stage, the validation set is leveraged for tuning the hyperparameters. After completing the initial four epochs training, an adaptive learning rate adjustment strategy is implemented to optimize the learning rate. The optimization process serves to fine-tune the learned model parameters so as to enhance the prediction performance of the model and reduce any potential overfitting or underfitting issues.

The final stage is the verification phase, where the test set is employed to evaluate the generalization ability of the proposed FEFD-YOLOV5 algorithm. This involves assessing the model's performance when it is exposed to unseen data, thereby testing its robustness

and generalization capacity. Moreover, to evaluate the performance of the algorithm under noise in the experiment, we chose a type of noise—Gaussian noise—for noise addition processing of the test images.

We have made three improvements on the YOLOv5s model, which include the addition of a Small Object Detection Head (SDH), attention modules selection, and integration of a denoising module. The effectiveness of these improvements was validated through a series of independent and integrated experiments.

To verify the effectiveness of attention and feature network fusion, SENet and CBAM, two different attention mechanisms, were added to the main network, and SDH was added to the detection layer for feature enhancement. The results of these experiments are depicted in Table 2.

Table 2. Ablation experiment results comparison.

SENet	CBAM	SDH	mAP@0.5/%	t/ms
			88.01%	24
✓			92.43%	27
	✓		89.63%	28
		✓	90.47%	26
	✓	✓	92.60%	32
✓		✓	94.78%	30

The results in Table 2 clearly demonstrate that in the network models that incorporate SENet and CBAM attention modules, the SENet module can significantly improve the mAP@0.5 index. The performance gain of the CBAM module on this test set is relatively smaller. By comparing the fusion experiments of SENet with SDH and CBAM with SDH, we found that SENet takes less time than CBAM, and outperforms CBAM in performance. Given the improvements in mAP@0.5 index achieved by our model modifications, we conducted an analysis. Based on the results of independent experiments, we calculated the average percentage improvement in mAP@0.5 for SDH, SENet, and CBAM. In addition, we also conducted an inferential statistical analysis by performing a *t*-test to determine the significance of the differences in mAP@0.5 improvements among SDH, SENet, and CBAM. Hence, the fusion of SENet attention mechanism with SDH was chosen as the optimization strategy for this research.

Although the introduction of SDH and SENet attention mechanism module increases the model's complexity, the number of parameters, and the computation load, thus reducing detection speed, the mAP@0.5 index improves by 6.77% on the original basis, which still meets the real-time performance requirement. Overall, the proposed model maintains good real-time performance while ensuring a high degree of accuracy.

Additionally, the experimental results show the advantages of SDH in small object detection. From Table 2, it can be observed that the mAP@0.5 improves by 2.46% when SDH is introduced alone. When fused with SENet or CBAM, the addition of SDH improves mAP@0.5 by 2.35% and 1.97%, respectively. These data suggest that SDH has a significant effect in enhancing the performance of small object detection. The detection effect is shown in Figure 7. Moreover, when SDH is fused with SENet and CBAM attention mechanisms in the experiments, the impact on detection speed is relatively small, with only an additional processing time of 2–4 ms. This further verifies the effectiveness and practicality of SDH.

In the fusion experiment, the combination of SENet and SDH performed the best, achieving an mAP@0.5 of 94.78% and a processing time of 30 ms, an improvement of 6.77% in accuracy compared to the original model, with only a sacrifice of 6 ms in detection speed. The detection results comparison is shown in Figure 8. This fully illustrates the effectiveness and synergistic effect of the SENet attention mechanism and SDH in improving model performance, allowing the model to maintain high accuracy while still having good real-time performance.



(**a**) YOLOv5s model (**b**) SDH

Figure 7. Comparison of detection results.



(**a**) YOLOv5s model (**b**) Attention module

Figure 8. Comparison of detection results.

This paper proposes a structure named FEFDNet, which integrates SDH, SENet, and a denoising module. To verify the effectiveness of the denoising module, we compared FEFDNet with YOLOv5s and YOLOv5s-N (a variant of YOLOv5s integrated with a denoising module). As can be seen from Table 3, FEFDNet achieves higher detection accuracy on datasets with no noise and different noise levels. The detection results comparison is shown in Figure 9.



(**a**) YOLOv5s model (**b**) Denoise module

Figure 9. Comparison of detection results.

Table 3. Comparison of detection results between FEFDNet and other algorithms.

Model	No Noise	σ = 15	σ = 25	σ = 35	σ = 50
YOLOv5s	90.26%	88.50%	85.65%	80.92%	70.17%
YOLOv5s-N	90.50%	91.02%	91.43%	90.78%	89.40%
FEFDNet	94.89%	94.33%	94.57%	93.82%	91.55%

On datasets with different noise levels, the detection accuracy of FEFDNet and YOLOv5s-N is generally higher than YOLOv5s. Especially on datasets with higher noise levels, the performance improvements of FEFDNet and YOLOv5s-N compared to YOLOv5s are particularly noticeable. For instance, when = 50, the detection accuracy of FEFDNet and YOLOv5s-N is 91.55% and 89.40%, respectively, while the detection accuracy of YOLOv5s is 70.17%. Compared to YOLOv5s, FEFDNet improves the detection accuracy on datasets with higher noise by 21.38 percentage points, and YOLOv5s-N improves by 19.23 percentage points. This evidence implies that the denoising module proposed herein can significantly enhance the model's robustness against noise data.

To better verify the model, this paper uses the same helmet dataset. Under the same configuration environment, two-stage detection model Faster R-CNN and single-stage detection models SSD, YOLOv3, YOLOv4, YOLOv5s, YOLOv6, and YOLOv7 were used to train helmet detection models. The comparison between these models and the FEFDNet

model was based mainly on AP@0.5 and mAP@0.5 as the key comparison indicators, and the number of epochs for training was 300. The comparison results of the experiments are shown in Table 4. The safety helmet detection model based on feature enhancement can detect targets wearing safety helmets with an AP of 94.89%. At the same time, this model can detect targets not wearing safety helmets, with an AP of 94.67%. Considering both, the mAP of this model is 94.78%. Compared with other algorithms, this model's mAP has improved significantly: Compared to Faster RCNN, SSD300, YOLOv3, YOLOv4, YOLOv5, YOLOv6, and YOLOv7, this model's mAP increased by 11.17%, 18.55%, 14.05%, 7.88%, 6.77%, 5.8%, and 5.07%, respectively. This indicates that the FEFDNet model has higher accuracy and performance advantages in construction site safety helmet detection.

Table 4. Comparison of experimental results of multiple models.

Algorithm	AP@0.5/% (Hat)	AP@0.5/% (Person)	mAP@0.5/%
Faster R-CNN	82.67%	84.55%	83.61%
SSD300	77.62%	74.84%	76.23%
YOLOv3	81.78%	79.68%	80.73%
YOLOv4	87.87%	85.93%	86.90%
YOLOv5s	90.26%	85.76%	88.01%
YOLOv6 [26]	90.46%	87.51%	88.98%
YOLOv7 [27]	90.67%	88.76%	89.71%
FEFDNet	94.89%	94.67%	94.78%

In conclusion, the improvement strategy based on the YOLOv5s model in this paper, which includes adding small target detection heads, selecting the optimal attention module, and integrating a denoising module, has proven to provide significant advantages. The experimental results show that FEFDNet has higher accuracy and robustness in the task of safety helmet detection, providing strong support for practical applications. Furthermore, in terms of performance, FEFDNet surpasses other mainstream object detection algorithms, further substantiating the efficacy of the improvements proposed in this study.

5. Conclusions

To meet the demands of helmet detection tasks in practical applications, this paper proposes an enhanced helmet detection algorithm based on the YOLOv5s model, integrating a small object detection head (SDH), optimally selected attention modules, and a denoising module. The efficacy and performance advantage of the proposed method are verified through a series of independent and combined experiments.

To address the small object detection problem, the SDH was introduced to improve the model's ability to detect small objects. In terms of attention mechanisms, the Squeeze-and-Excitation Networks (SENet) was chosen over Convolutional Block Attention Module (CBAM) as it proved to be more effective. To enhance the robustness of the model under noisy data, a denoising module was designed and integrated into the FEFDNet network.

To evaluate the performance of the proposed method, FEFDNet was compared with other mainstream object detection algorithms. The experimental results indicate that FEFDNet outperforms other algorithms in helmet detection tasks, demonstrating good robustness under different noise conditions, lighting conditions, and degrees of occlusion. This fully illustrates the high value of the proposed method in practical applications.

Although the proposed FEFDNet has achieved commendable performance, there is still room for improvement. For instance, further research into more effective small object detection and tracking algorithms could improve the recognition accuracy of these objects. Additionally, exploration of how the proposed method could be applied to other object detection algorithms to achieve superior performance can be considered.

Author Contributions: Conceptualization, Y.Z. and H.B.; methodology, Y.Q.; software, Y.Z. and Y.Q.; validation, Y.Z. and Y.Q.; formal analysis, Y.Z. and Y.Q.; investigation, Y.Z. and Y.Q.; resources, Y.Z. and Y.Q.; data curation, Y.Z. and Y.Q.; writing—original draft preparation, Y.Z. and H.B.; writing—review and editing, Y.Z. and Y.Q.; visualization, Y.Z. and H.B.; supervision, Y.Z. and H.B.; project administration, Y.Z. and H.B.; funding acquisition, Y.Z. and H.B. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported in part by National Key R & D Program of China (No. 2022YFE0200300), National Natural Science Foundation of China (No. 61972023), Beijing Natural Science Foundation (L223022) .

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
2. Redmon, J.; Farhadi, A. YOLO9000: Better, faster, stronger. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7263–7271.
3. Redmon, J.; Farhadi, A. Yolov3: An incremental improvement. *arXiv* **2018**, arXiv:1804.02767.
4. Bochkovskiy, A.; Wang, C.Y.; Liao, H.Y.M. Yolov4: Optimal speed and accuracy of object detection. *arXiv* **2020**, arXiv:2004.10934.
5. Jeong, J.; Park, H.; Kwak, N. Enhancement of SSD by concatenating feature maps for object detection. *arXiv* **2017**, arXiv:1705.09587.
6. Li, Z.; Zhou, F. FSSD: Feature fusion single shot multibox detector. *arXiv* **2017**, arXiv:1712.00960.
7. Fu, C.Y.; Liu, W.; Ranga, A.; Tyagi, A.; Berg, A.C. Dssd: Deconvolutional single shot detector. *arXiv* **2017**, arXiv:1701.06659.
8. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
9. Girshick, R.; Donahue, J.; Darrell, T.; Malik, J. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 580–587.
10. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
11. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks. *Adv. Neural Inf. Process. Syst.* **2015**, *28*, 1–9. [CrossRef] [PubMed]
12. He, K.; Gkioxari, G.; Dollár, P.; Girshick, R. Mask r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2961–2969.
13. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
14. Liu, S.; Qi, L.; Qin, H.; Shi, J.; Jia, J. Path aggregation network for instance segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8759–8768.
15. Chen, P.Y.; Chang, M.C.; Hsieh, J.W.; Chen, Y.S. Parallel residual bi-fusion feature pyramid network for accurate single-shot object detection. *IEEE Trans. Image Process.* **2021**, *30*, 9099–9111. [CrossRef] [PubMed]
16. Liu, S.; Huang, D. Receptive field block net for accurate and fast object detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 385–400.
17. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
18. Woo, S.; Park, J.; Lee, J.Y.; Kweon, I.S. Cbam: Convolutional block attention module. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 3–19.
19. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.
20. Han, K.; Zeng, X. Deep learning-based workers safety helmet wearing detection on construction sites using multi-scale features. *IEEE Access* **2021**, *10*, 718–729. [CrossRef]
21. Li, N.; Lyu, X.; Xu, S.; Wang, Y.; Wang, Y.; Gu, Y. Incorporate online hard example mining and multi-part combination into automatic safety helmet wearing detection. *IEEE Access* **2020**, *9*, 139536–139543. [CrossRef]
22. Wu, F.; Jin, G.; Gao, M.; Zhiwei, H.; Yang, Y. Helmet detection based on improved YOLO V3 deep model. In Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control (ICNSC), Banff, AB, Canada, 9–11 May 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 363–368.
23. Han, G.; Zhu, M.; Zhao, X.; Gao, H. Method based on the cross-layer attention mechanism and multiscale perception for safety helmet-wearing detection. *Comput. Electr. Eng.* **2021**, *95*, 107458. [CrossRef]

24. Tai, W.; Wang, Z.; Li, W.; Cheng, J.; Hong, X. DAAM-YOLOV5: A Helmet Detection Algorithm Combined with Dynamic Anchor Box and Attention Mechanism. *Electronics* **2023**, *12*, 2094. [CrossRef]
25. Murtagh, F. Multilayer perceptrons for classification and regression. *Neurocomputing* **1991**, *2*, 183–197. [CrossRef]
26. Li, C.; Li, L.; Jiang, H.; Weng, K.; Geng, Y.; Li, L.; Ke, Z.; Li, Q.; Cheng, M.; Nie, W.; et al. YOLOv6: A single-stage object detection framework for industrial applications. *arXiv* **2022**, arXiv:2209.02976.
27. Wang, C.Y.; Bochkovskiy, A.; Liao, H.Y.M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 7464–7475.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.