Renewable Energy and Power Quality Journal
https://repqj.com/index.php/repqj/index
RE&PQJ, Vol. 23, No. 5, 2025

Method for Identifying Dangerous Operation Actions at Power Marketing and Energy Metering Sites Using the Mask R-CNN Image Segmentation Model

Lei Wei[1,*], Liang Wang[1], Feihong Yin[1], Junchen Guo[2]

[1]NARI Nanjing Control System Co. Ltd, Nanjing, 211106, Jiangsu, China
[2]Hohai University, Changzhou, 213000, Jiangsu, China
*Corresponding author email: dingj_01@163.com

Abstract. Safety issues at power marketing electricity metering sites are related to the personal safety of staff and the stable operation of the power system. Accurate identification of dangerous operating actions is crucial.Traditional CNN recognition is easily affected by background interference in recognizing dangerous operation actions at power marketing and energy metering sites and has a weak ability to express important features. This paper uses the improved Mask R-CNN image segmentation model to perform pixel-level segmentation on dangerous operation actions at the power marketing and energy metering site, and accurately locates its boundaries and actions. The study first uses the GAN model to expand the dangerous operation action images to ensure data balance, and uses the ResNeXt module to replace the ResNet in the traditional Mask R-CNN for feature extraction. Then, CBAM is embedded in the feature extraction module to enhance the extraction of spatial and temporal information of dangerous operation actions and reduce background interference. Finally, the loss function is optimized by combining Boundary loss to reduce the impact of the missing edge of the dangerous operation mask. This paper conducts experiments based on images of an actual power work site from June to December 2024. The results show that the improved Mask R-CNN performs best, with an accuracy of 96.9%, which is 2.6% higher than Mask R-CNN, and MIoU reaches 95.9%. The experimental results show that combining the Mask R-CNN image segmentation model and optimization module can effectively improve the accuracy and segmentation accuracy of dangerous operation action recognition at the power marketing and energy metering site, and ensure the safety of operators.

Key words. Electricity marketing energy metering site, Dangerous operation, Action recognition, Mask R-CNN model, ResNeXt module, CBAM module

1. Introduction

The development of smart grids has made electricity marketing and metering crucial in ensuring supply and demand balance, optimizing resource allocation and improving user experience [1-3]. In actual power marketing and electric energy metering sites, workers often face complex operating environments and high-risk work tasks, such as dangerous operations such as contact with high-voltage wires, equipment failure handling, and high-altitude operations [4]. If dangerous operations on site are not discovered and corrected in time, serious safety accidents can occur, threatening the lives of workers and the stable operation of the power system [5,6]. At present, the action recognition of models such as CNN (Convolutional Neural Networks) is difficult to resist background interference, and the ability to represent important features is not strong, resulting in poor action recognition accuracy.

In recent years, researchers have used computer vision and deep learning techniques to identify operational actions in fields such as electricity and solved some challenges. Xin, Helmi and other scholars combined skeletons and used Transformer, RCNN-BiGRU (Region-Convolutional Neural Networks-Bidirectional Gated Recurrent Unit) for action recognition, which improved the recognition accuracy [7,8]. Yang and other scholars combined GCN (Graph Convolutional Network) and CNN to carry out action recognition of operators and solved the dependency problem between different joints of the human body [9]. The CNN model, YOLO-v5, has been widely used in the motion detection of operators on construction sites. It optimizes the motion recognition performance, ensures the safety of operators, and provides a reference for future scholars in other fields [10,11]. Park et al. applied the YOLO (You Only Look Once) model to worker detection at construction sites,

achieving an accuracy of 77.57% [12]. CNN-LSTM (Convolutional Neural Networks-Long Short-Term Memory), as a hybrid model, achieved an accuracy of 90.89% in human activity recognition, optimizing overall recognition performance [13,14]. YOLO performs well in real-time target detection, but its fine-grained feature extraction capability is weak in complex backgrounds, making it difficult to accurately segment the boundaries of dangerous operations. CNN-based methods have achieved a certain degree of accuracy improvement in action recognition, but their ability to suppress background interference is insufficient, making it difficult to extract key action features.Scholars used CNN and other models to identify operators' actions, which improved the accuracy of action recognition to a certain extent, but the performance was weak in resisting background interference, and the ability to express important features in the action was lacking.

In the power industry, research not only focuses on operator safety and motion recognition, but also involves power demand forecasting and electricity price analysis to optimize the operation of the power system. Shah I and other scholars compared different modeling methods to improve the accuracy of power demand and price forecasting [15]. Iftikhar H and other scholars proposed a novel homogeneous and heterogeneous integrated learning method to improve the accuracy of power consumption forecasting [16]. Gonzales S M and others used an improved time series integration method to analyze and predict electricity prices, and achieved good application results in the Peruvian power market [17]. At the same time, Iftikhar H and others further studied power demand forecasting based on time series integration technology, providing data support for intelligent dispatching and safety management in the power industry [18]. These research results are not only of great value in power dispatching and market analysis, but also provide new ideas for intelligent monitoring and safety prevention and control in future power operation scenarios.

To solve the current problem, more and more studies have begun to introduce image segmentation models to improve the recognition accuracy of dangerous operation actions. Li et al. used Faster R-CNN to detect workers' construction activities and identify their action states, with a significant improvement in average accuracy [19,20]. Mask R-CNN, as an improved CNN, adds pixel-level segmentation capabilities on the basis of target detection. It is used in action recognition to effectively solve the background interference problem and can accurately locate the action boundary [21,22]. DHIVYA and other scholars applied Mask R-CNN to target detection in surveillance videos in construction and other fields, improving the accuracy of target detection [23]. The Mask R-CNN model has many applications in action recognition, greatly improving the fine segmentation of edges and action recognition performance [24-26]. Scholars applied the Mask R-CNN model to action recognition, which improved the anti-interference ability of the background compared to CNN, but still failed to solve the problem of expressing important features. In addition, there are few studies on the Mask R-CNN model in the recognition of dangerous operation actions in power marketing and energy metering, and there are many research gaps.

This study aims to solve the key problem of identifying dangerous operation actions in power marketing and energy metering, and how to improve the recognition accuracy and segmentation effect of traditional methods under complex backgrounds. This paper adopts the improved Mask R-CNN (Mask Region-based Convolutional Neural Network) image segmentation model, and improves the performance of the model in feature extraction, background suppression and boundary accuracy by introducing a series of optimization modules. The experiment replaced the traditional ResNet (Residual Network) with the ResNeXt (Residual Networks with Next) module, and introduced the CBAM (Convolutional Block Attention Module) module to further reduce background interference. Finally, the boundary segmentation accuracy was optimized by combining the Boundary loss function. In the actual power site image test, the recognition accuracy and segmentation precision were 96.9% and 95.9% respectively, which were significantly improved. This proves that the experimental method has high application potential in improving the effect and practicality of identifying safe operations at power sites.

Contribution of the paper:

(1) This paper replaces the traditional ResNet with the ResNeXt module, and combines the CBAM module and the Boundary loss function to significantly improve the accuracy of feature extraction, background suppression and boundary segmentation, overcoming the recognition bottleneck of the traditional CNN model in complex power field environments.

(2) The experiment uses the GAN (Generative Adversarial Network) model to expand the dangerous operation action images to solve the data imbalance problem, ensure the stability and efficiency of model training, and enhance the model's generalization ability for different scenarios.

(3) The study tested the image data of the actual power marketing power metering site and achieved significant recognition results, proving the practical application potential of the improved Mask R-CNN model in improving the recognition accuracy and segmentation effect of dangerous operation actions at the power site.

The structure of this paper is as follows:

Chapter 1 is the introduction, which explains the background and research significance of the topic, the purpose of the research, the hypothesized questions, and the contribution and structure of the text.

Chapter 2 is the method part, which discusses the experimental data and data preprocessing, and introduces how to use GAN to expand the data set. It explains how to improve the Mask R-CNN model, the principle of CBAM, and explains the training and optimization strategies of the experiment.

Chapter 3 is the results and discussion. This chapter explains the evaluation indicators and experimental design, verifies the experimental results from the aspects of recognition accuracy and image segmentation accuracy, and discusses the reasons behind the excellent methods of the article, the impact and significance of the article, and the discussion of the limitations and future directions of the article.

Chapter 4 is the conclusion part, which summarizes the main achievements of the article and points out the future research direction.

2. Research Methods

A. Experimental Data

The data for the tests in this paper come from images taken at actual power engineering sites, covering images of dangerous operations such as failure to wear a safety helmet, failure to wear insulating gloves, failure to wear a safety belt, accidental contact with live tools, irregular operation of exposed wires, and high-voltage operation without cutting off the power supply. The data collection time is from June to December 2024, and a total of 5268 images were collected, including normal operations and dangerous operations. The experiment uses ten-fold cross validation to divide the data, and finally removes the mean as the final result. Some experimental images are shown in Figure 1.

Figure 1. Some experimental images.

The data is very rich in Figure 1, including images of various dangerous operations in the power marketing and energy metering site.

B. Data Preprocessing

1) Denoising and Image Normalization

The image data of the actual power work site has some background noise. This paper uses Gaussian filtering technology [27,28] to denoise the image.

In order to facilitate the model processing, this paper uses the minimum-maximum normalization to normalize each pixel value to the range of [0, 1].

2) Image Annotation

This paper uses the annotation tool LabelMe to perform rectangular annotation on dangerous operation areas in the collected images, and for the image segmentation task, the experiment uses pixel-level masks for annotation. Each dangerous operation action area is represented by a binary image, where the target area is 1 and the background area is 0.

C. GAN Model Data Expansion

The number of samples of dangerous operation action images at the power marketing and energy metering site is relatively small. Directly using the original data set can cause the model to tend to predict the more common normal operations and ignore the identification of dangerous operations. This paper introduces the GAN model [29-31] to expand the data of dangerous operation action images.

Reference [29] discussed the application of GAN in text data enhancement, mainly for natural language processing, but its data generation strategy is of reference significance for the expansion of image data. Reference [30] proposed a GAN-based data enhancement method for forest mapping, which improves the generalization ability of the model by balancing the data distribution, which is similar to the goal of balancing the dangerous operation data set through GAN in this study. Reference [31] combined GAN with CNN-LSTM for aerial activity

recognition, proving the ability of GAN to generate high-quality data in complex backgrounds, which is consistent with the idea of improving dangerous operation recognition in this study. Studies have shown that GAN can effectively expand the data set, improve the model's recognition ability for minority class samples, and improve the robustness and generalization ability of the model.

The generator and the discriminator are trained by mutual game, and finally the generator can generate high-quality and real image data.

For data augmentation, first define a random noise vector. The generator samples from the noise space and inputs it into the generator network. The generator generates an image based on the input noise vector. The optimization goal of the generator network is to maximize the error of the discriminator, which corresponds to the discriminator considering the generated image to be a real image. The loss function of the generator is shown in formula (1).

$$\text{Loss}_G = -E_{\alpha \sim q_\alpha(\alpha)}\left[\log D\left(G\left(\alpha\right)\right)\right] \quad (1)$$

D represents the discriminator network, E represents expectation, $q_\alpha(\alpha)$ represents the probability distribution of noise. $G(\alpha)$ represents the output of the generator.

The task of the discriminator is to classify the input image and determine whether it comes from the real data distribution. The loss function of the discriminator is shown in formula (2).

$$\text{Loss}_D = -E_{p \sim q_{td}(p)}\left[\log D(p)\right] - E_{\alpha \sim q_\alpha(\alpha)}\left[\log\left(1 - D(G(\alpha))\right)\right] \quad (2)$$

$q_{td}(p)$ represents the real data distribution. $D(p)$ represents the output value of the discriminator for the image.

GAN is trained by alternately optimizing the generator and the discriminator. In each training step, the parameters of the discriminator are first updated to distinguish the real image from the generated image as much as possible. The parameters of the generator can be updated so that the images it generates are judged as real images by the discriminator as much as possible. Finally, multiple iterations of optimization are performed to enable the generator to generate high-quality, realistic images of dangerous operation actions.

The comparison of data of each category before and after expansion is shown in Table 1.

Table 1. Comparison of data of each category before and after expansion

Types	Before expansion	Types	After GAN expansion
Normal and standardized operation	3074	Normal and standardized operation	3074
Failure to wear a safety helmet before live work	301	Failure to wear a safety helmet before live work	521
Not wearing insulating gloves	479	Not wearing insulating gloves	503
Not wearing a safety belt	433	Not wearing a safety belt	534
Inadvertent contact with live tools	174	Inadvertent contact with live tools	469
Irregular operation of exposed wires	259	Irregular operation of exposed wires	458
High voltage operation without cutting off the power supply	548	High voltage operation without cutting off the power supply	590

In Table 1, after the GAN model is expanded, the action categories of failure to wear a helmet before live work, accidental contact with live tools, and irregular operation of exposed wires are most significantly expanded. The overall ratio of dangerous operations to normal operations is close to 1, and the types of dangerous operations are relatively balanced without obvious differences, ensuring the feasibility of the experiment.

D. Improved Mask R-CNN Image Segmentation Model

1) ResNeXt Module

This paper uses the ResNeXt module [32,33] to replace the traditional ResNet module for feature extraction. The ResNeXt module has significant advantages over the ResNet network. Without obvious changes in the magnitude of parameters, the three-layer convolutional blocks of the original ResNet network are replaced by parallel stacking of blocks with the same topology structure to ensure the accuracy of feature extraction.

ResNeXt replaces a single convolution with a multi-branch convolution based on ResNet. In this experiment, the dangerous operation action images input at the power marketing and energy metering site are sent to each branch for convolution operations, and the feature maps output by each branch are spliced and the extraction results are output.

This paper chooses ResNeXt as the feature extraction network, mainly based on its balance between computational efficiency and feature expression ability. Compared with ResNet-50, ResNeXt adopts a multi-branch group convolution structure, which is

expected to improve the model's feature extraction and expression capabilities without significantly increasing the number of parameters, making it more robust in the recognition of dangerous operation actions in complex power marketing and energy metering sites. ResNeXt has better parallel computing capabilities than EfficientNet, and its group convolution structure can effectively reduce computational redundancy, improve feature diversity, and is more suitable for pixel-level segmentation tasks. EfficientNet mainly relies on a compound scaling strategy for model optimization, which is feasible for image classification tasks, but not for target detection and segmentation tasks. This paper uses ResNeXt as a feature extraction network to ensure that the model's recognition ability for dangerous operation actions is guaranteed while the computational cost is controllable.

In ResNeXt, the convolution operation adopts the form of group convolution, and the convolution operation of each branch is performed on different feature subsets. The convolution operation in the ResNeXt module is expressed as shown in formula (3).

$$f_i = \mathrm{Conv}\left(Z_i, U_i\right) + \beta_i \quad (3)$$

Z_i represents the input image.

ResNeXt introduces the Carlson and group convolution mechanisms to decompose each convolution into several small convolution operations.

The convolution operation performed in each group is expressed as shown in formula (4).

$$f_i = \mathrm{Conv}_{\gamma}\left(Z_i, U_i\right) + \beta_i \quad (4)$$

γ represents the number of groups.

In the ResNeXt module, each convolution layer performs a single convolution operation, while multiple convolution units with the same structure work in parallel. Each convolution block includes multiple parallel branches, and the input image is convolved through each branch. The output feature map can be spliced along the channel dimension, and the final feature map is expressed as shown in formula (5).

$$F = \left[f_1, f_2, \cdots, f_{\gamma} \right] \quad (5)$$

f_{γ} represents the convolution output of the branch.

2) CBAM Module

In the image recognition task of dangerous operation actions in the power marketing and energy metering field, the image background often interferes with the recognition of the target, especially in complex scenes, the features of dangerous operation actions are often very small and easily affected by noise. This paper introduces the CBAM module [34,35] in the improved Mask R-CNN framework. By guiding the attention mechanism in the spatial and channel dimensions, the feature expression ability of dangerous operation actions in space and channels is enhanced. In the CBAM module [36], it consists of two parts: the Channel Attention Mechanism (CAM) and the Spatial Attention Mechanism (SAM).

(1) CAM

CAM enhances the feature expression of the target area by assigning different attention weights to each position of the input feature map. The experiment uses two operations, Global Average Pooling (GAP) and Global Max Pooling (GMP), to generate channel-level attention weights.

For the input feature map, this paper first performs global pooling on each channel to generate two channel descriptors, namely GAP and GMP. The calculation formulas of GAP and GMP are shown in formulas (6) and (7).

$$M_{\mathrm{avg}}\left(F\right) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} F\left(i, j, :\right) \quad (6)$$

$$M_{\max}\left(F\right) = \max_{i,j} F\left(i, j, :\right) \quad (7)$$

F represents the input feature map. $M_{\mathrm{avg}}\left(F\right)$ represents the calculation result of GAP, and $M_{\max}\left(F\right)$ represents the calculation result of GMP.

After obtaining the descriptors of the two channels, the experiment passes them into a shared fully connected layer and transforms them using the ReLU activation function. The calculation formulas for the transformation are shown in formula (8) and formula (9) respectively.

$$g_{\mathrm{avg}}\left(F\right) = \mathrm{ReLU}\left(W_{\mathrm{avg}} \cdot M_{\mathrm{avg}}\left(F\right) + \delta_{\mathrm{avg}}\right) \quad (8)$$

$$g_{\max}\left(F\right) = \mathrm{ReLU}\left(W_{\max} \cdot M_{\max}\left(F\right) + \delta_{\max}\right) \quad (9)$$

W_{avg} and $W_{\max}$ represent the weights of the fully connected layer, δ_{avg} and $\delta_{\max}$ are bias terms.

The experiment combines the two transformed features and uses a Sigmoid activation function to generate the channel attention (CA) weight. The expression of CA weight is shown in formula (10).

$$B_c(F)=\sigma\left(g_{avg}(F)+g_{max}(F)\right) \quad (10)$$

$B_c(F)$ represents the attention weight of each channel.

For the input feature map, it is weighted according to the channel, as shown in the formula (11).

$$F'=F\times B_c(F) \quad (11)$$

F' represents the weighted feature map.

(2) SAM

This paper first sums the feature map weighted by CA along the channel dimension to generate two different spatial descriptors, namely average pooling and maximum pooling. The calculation formulas of average pooling and maximum pooling are shown in formulas (12) and (13).

$$M^s_{avg}(F')=\frac{1}{C}\sum_c^C F'(i,j,c) \quad (12)$$

$$M^s_{max}(F')=\max_c F'(i,j,c) \quad (13)$$

$M^s_{avg}(F')$ represents the result of average pooling, and $M^s_{max}(F')$ represents the result of maximum pooling.

The two descriptors $M^s_{avg}(F')$ and $M^s_{max}(F')$ are input into a shared convolutional layer, and the spatial attention (SA) weight map is generated through the Sigmoid activation function. After the two spatial descriptors are output through the shared convolutional layer, the calculation is shown in formulas (14) and (15).

$$g^s_{avg}(F')=\mathrm{Conv}\left(M^s_{avg}(F'),W^s_{avg}\right) \quad (14)$$

$$g^s_{max}(F')=\mathrm{Conv}\left(M^s_{max}(F'),W^s_{max}\right) \quad (15)$$

W^s_{avg} and W^s_{max} represent convolution kernels.

The two descriptors $g^s_{avg}(F')$ and $g^s_{max}(F')$ are combined by sum operation and input into the Sigmoid activation function to generate SA weights. The expression of SA weights is shown in formula (16).

$$B_s(F')=\sigma\left(g^s_{avg}(F')+g^s_{max}(F')\right) \quad (16)$$

$B_s(F')$ represents the generated SA weight.

Now apply the SA weight to the weighted feature map and output the spatially weighted feature map. The calculation is shown in formula (17).

$$F''=F'\times B_s(F') \quad (17)$$

F'' represents the feature map after SAM weighting.

The improved Mask R-CNN model is shown in Figure 2.

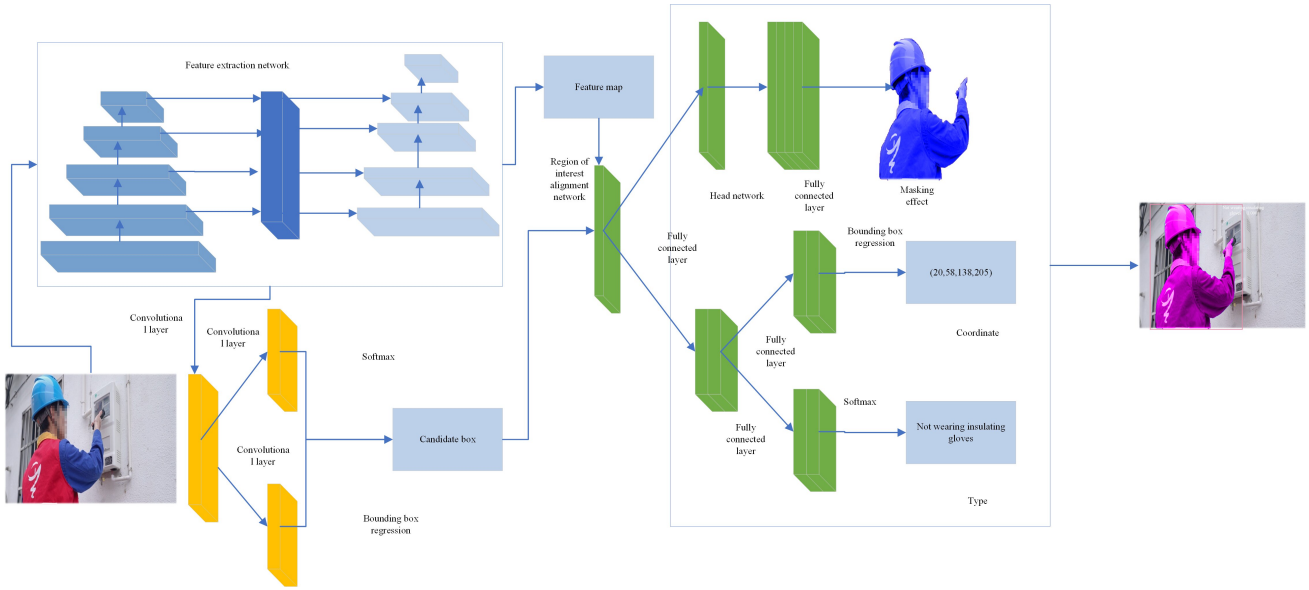

Figure 2. Improved Mask R-CNN model diagram.

In Figure 2, the improved Mask R-CNN model introduces the CBAM module and replaces the ResNet network with the ResNeXt module.After the image in Figure 2 is processed by the Mask R-CNN model, it can output dangerous operation actions and bounding box regression results well, and the effect of construction

personnel segmentation is obvious.

3) *Boundary Loss Function*

This paper introduces the Boundary loss function to optimize the segmentation mask loss. The Boundary loss function uses the boundary matching degree to supervise the loss function of the network. The pixels that match the boundary between the image and the real border are marked as 0, and the loss function is evaluated based on the distance from the border for the pixels that do not match the boundary.

The boundary loss is calculated by integrating the boundary. The boundary loss function is expressed as shown in formulas (18) and (19).

$$D_i(\partial R, \partial O) = 2\left(\int \eta_R(p)\, o(p)\, \mathrm{d}p - \int \eta_R(p)\, r(p)\, \mathrm{d}p\right) \quad (18)$$

R represents the area of the real box, and O represents the predicted area to be segmented. ∂R represents the boundary of the true box, and ∂O represents the boundary of the predicted area.

$$\mathrm{Loss}_{BD} = \int \eta_R(p) O_\zeta(p) \mathrm{d}p \quad (19)$$

$O_\zeta(p)$ represents the probability output of the network.

$\eta_R(p)$ represents the boundary level set.

An example of image visualization after the introduction of the Boundary loss function is shown in Figure 3.

Figure 3. Image visualization example after introducing the boundary loss function.

In Figure 3, the first picture is the original image, and the second picture is the segmentation result without the boundary loss. It can be observed that there is an obvious lack of boundaries, and the edges of some dangerous operation areas are blurred; the third picture is the segmentation result after the boundary loss is introduced. Compared with the second picture, its boundaries are clearer and the contours are complete, and it can match the real area more accurately. This shows that the boundary loss function can significantly reduce the discontinuity of the mask boundary, improve the segmentation accuracy, and optimize the recognition effect of dangerous operation actions by strengthening the boundary matching degree.

4) *Training and Optimization of the Improved Mask R-CNN Model*

The loss function of the Mask R-CNN model, the total loss function is expressed as shown in formula (20).

$$\mathrm{Loss} = \mathrm{Loss}_c + \mathrm{Loss}_b + \mathrm{Loss}_m \quad (20)$$

Loss_c represents classification loss, Loss_b represents bounding box regression loss, and Loss_m represents segmentation mask loss.

The classification loss is expressed as shown in formula (21).

$$\mathrm{Loss}_c = -\sum_i \lambda_i \log\left(\hat{\lambda}_i\right) \quad (21)$$

The bounding box regression loss as shown in formula (22).

$$\mathrm{Loss}_b = \sum_i \mathrm{smooth}\ L1\left(\varpi_i - \hat{\varpi}_i\right) \quad (22)$$

The segmentation loss is expressed as shown in formula (23).

$$\mathrm{Loss}_m = -\sum_{i,j} \lambda_{i,j} \log\left(\hat{\lambda}_{i,j}\right) + \left(1-\lambda_{i,j}\right)\log\left(1-\hat{\lambda}_{i,j}\right) \quad (23)$$

This study uses the Adam optimization algorithm [37-39] to adaptively adjust the learning rate, and it is shown in formula (24).

$$\rho_t = \rho_{t-1} - \iota \frac{\hat{\theta}_t}{\sqrt{\hat{\vartheta}_t + \epsilon}} \quad (24)$$

ϵ represents a tiny constant to prevent division by zero.

The hyperparameter settings of the improved Mask R-CNN model are shown in Table 2.

Table 2. Hyperparameters of the improved Mask R-CNN model

Parameters	Value	Parameters	Value
Learning rate	0.0001	Momentum factor	0.9
Number of iterations	300	Anchor frame	(16,32,64,128,256)
Confidence	0.7	Batch size	50
Optimizer	Adam	-	

In Table 2, the learning rate is 0.0001, the anchor box is (16, 32, 64, 128, 256). The batch size is 50, and the optimizer uses Adam.

3. Results and Discussion

A. Evaluation Indicators

mAP (Mean Average Precision):

$$mAP = \frac{1}{V}\sum_{v=1}^{V} AP_v \quad (25)$$

v represents the number of categories.

$MIoU$ (Mean Intersection over Union):

$$MIoU = \frac{1}{V}\sum_{v=1}^{V} IoU_v \quad (26)$$

$Accuracy$:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (27)$$

TP (True Positive) represents a true positive example, TN (True Negative) represents a true negative example. FP (False Positive) represents a false positive example, and FN (False Negative) represents a false negative example.

$Precision$:

$$Precision = \frac{TP}{TP + FP} \quad (28)$$

$Recall$:

$$Recall = \frac{TP}{TP + FN} \quad (29)$$

$F1$:

$$F1 = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (30)$$

B. Experimental Design

The experiment is divided into two aspects: action recognition performance verification and detection and segmentation performance. This paper sets up an experimental group and a control group. In the action recognition performance verification, the control models include DenseNet (Dense Convolutional Network), VGG 16 (Visual Geometry Group 16), Faster R-CNN (Faster Region-based Convolutional Neural Networks), and Mask R-CNN. In the detection and segmentation performance, the control models include SSD (Single Shot MultiBox Detector), YOLO v4, YOLO v8, and Mask R-CNN.

The study set up an ablation experiment to explore the impact of CBAM, ResNeXt module, and Boundary loss function on the performance of the Mask R-CNN model.

C. Dangerous Operation Action Recognition Performance

1) Dangerous Operation Action Recognition Result Display

The outcome of the identification of the hazardous operation actions is shown in Figure 4.

Figure 4. Dangerous operation action identification results.

In Figure 4, all kinds of dangerous operations at the power marketing and energy metering site can be well identified. The first picture was accurately identified as a dangerous operation without wearing insulating gloves, and the second picture was accurately identified as a high-voltage operation without cutting off the power supply. The third picture was accurately identified as an

irregular operation with exposed wires, and the fourth picture was accurately identified as an operation without wearing a safety helmet before live work.

2) *Confusion Matrix*

In the test set of this paper, there are 307 samples of normal and standardized operation, 52 samples of failure to wear a helmet before live work, and 50 samples of not wearing insulating gloves. There are 53 samples of not wearing a safety belt, and 47 samples of accidental contact with live tools. There are 46 samples of irregular operation of exposed wires, and 59 samples of high-voltage operation without cutting off the power supply. The confusion matrix is illustrated in Figure 5.

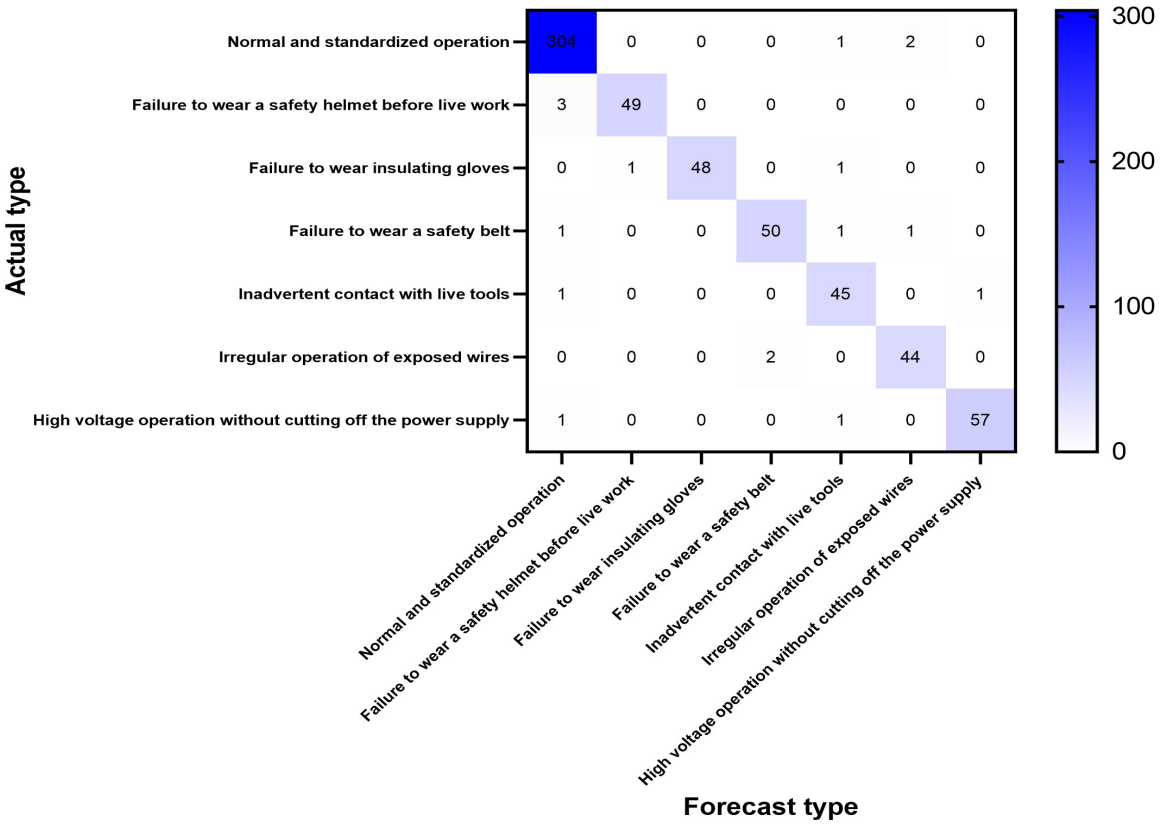

Figure 5. Confusion Matrix.

In Figure 5, 49 samples were correctly predicted to be the action of not wearing a helmet before live work, and 3 samples were incorrectly predicted as normal and standardized operation types. There were 304 samples of normal and standardized operation correctly predicted, 48 samples of operation without wearing insulating gloves correctly predicted, and 50 samples of operation without wearing safety belt correctly predicted. There were 45 samples of accidental touch of live tools correctly predicted, and 44 samples of improper operation of exposed wires correctly predicted. For high-voltage operation without cutting off the power supply, 57 samples were correctly predicted. Overall, the improved Mask R-CNN in this paper achieves good performance in identifying dangerous operations at the power marketing energy metering site.

3) *Dangerous Operation Action Recognition Performance Under Different Models*

The dangerous operation action recognition performance under different models is shown in Figure 6.

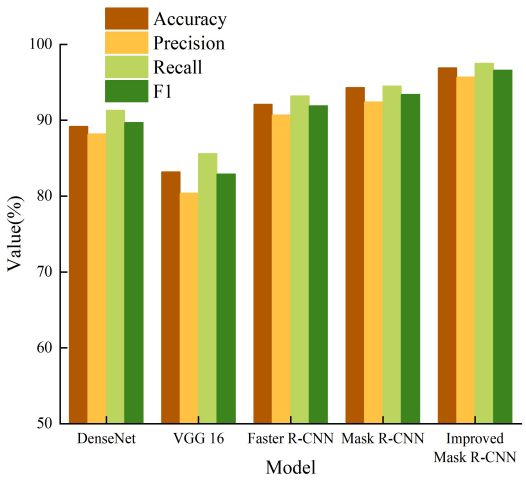

Figure 6. Dangerous operation action recognition performance under different models.

In Figure 6, the improved Mask R-CNN performs best, with an accuracy of 96.9% and an F1 value of 96.6%,

significantly higher than other models. Compared with Mask R-CNN, they have increased by 2.6% and 3.2% respectively. DenseNet has an accuracy of 89.2% and an F1 value of 89.7%. VGG 16 has a lower accuracy and F1 value of only 83.2% and 82.9%. Faster R-CNN has an accuracy of 92.1% and an F1 value of 91.9%.

The improved Mask R-CNN leads all models with a precision of 95.7% and a recall of 97.5%, both higher than other models. DenseNet has a precision of 88.2% and a recall of 91.3%. The improved Mask R-CNN introduces stronger feature extraction and enhancement modules to improve the accurate recognition of small targets and boundaries, effectively improving the recognition performance.

4) Detection and Segmentation Performance

The detection and segmentation performance is illustrated in Figure 7.

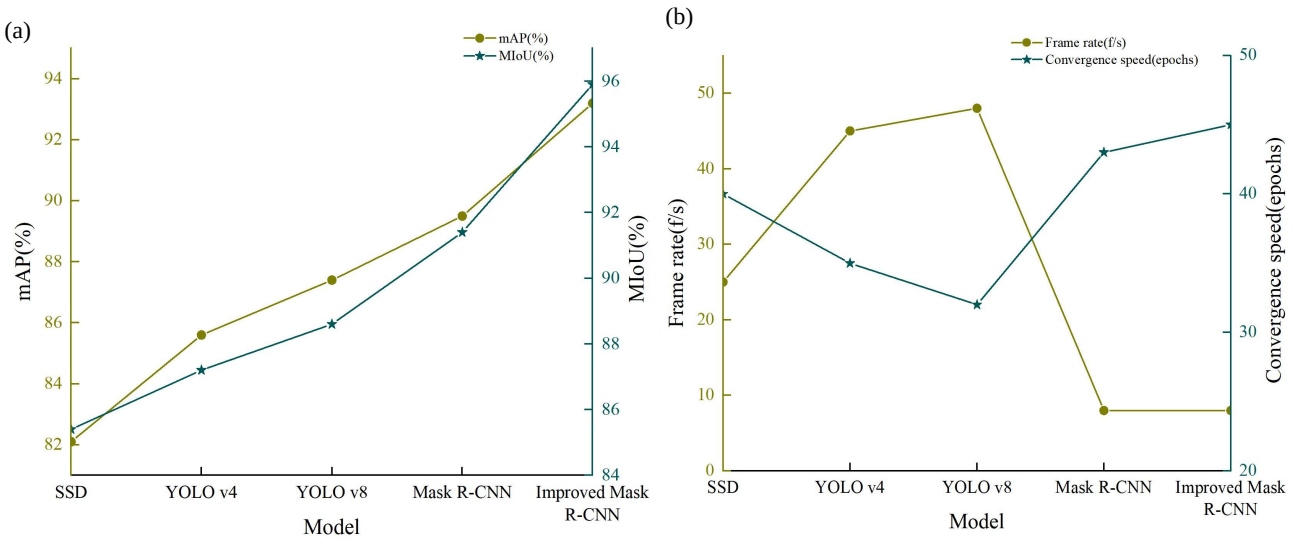

Figure 7. Detection and segmentation performance. Figure 7 (a) mAP and MIoU performance; Figure 7 (b) Frame rate and convergence speed performance.

In Figure 7 (a), the improved Mask R-CNN achieves 93.2% mAP and 95.9% MIoU, which is the most outstanding. Mask R-CNN has an mAP of 89.5% and an MIoU of 91.4%. YOLO v8 has mAP and MIoU of 87.4% and 88.6% respectively, while YOLO v4 and SSD perform relatively poorly.

In Figure 7(b), for frame rate, YOLO v8 performs best at 48 frames/second, while YOLO v4 reaches 45 frames/second. Mask R-CNN and Improved Mask R-CNN have lower frame rates of 8 frames/second. Mask R-CNN and its improved versions have a large amount of computation when performing image segmentation, need to process more image areas and details, and have a low frame rate, while the YOLO series and SSD focus on target detection and have relatively small computational workloads.

In terms of convergence speed, the convergence speed of the improved Mask R-CNN is 45 epochs, while Mask R-CNN only needs 43 epochs to converge.

5) Ablation Experiment

The outcome of the ablation test is illustrated in Table 3. In Table 3, * indicates Mask R-CNN.

Table 3. Ablation experiment results (%)

Model	*Accuracy*	*Precision*	*Recall*	*F*1
* (ResNet)	94.3	92.4	94.5	93.4
Improved * (ResNeXt)	94.9	93	95	94.0
Improved * (Boundary)	94.6	92.9	95.0	93.9
Improved * (CBAM)	95.5	94	95.8	94.9
Improved * (ResNeXt+CBAM)	96.1	94.7	96.2	95.4
Improved *(ResNeXt+CBAM+Boundary)	96.9	95.7	97.5	96.6

In Table 3, the improved masked R-CNN (ResNeXt+CBAM+Boundary) performs the best among all the metrics with 96.9% accuracy and 96.6% F1 value. Removing the boundary module decreases the accuracy to 96.1% and the F1 value to 95.4%. When the ResNeXt module is further removed, the accuracy decreases to

95.5% and the F1 value decreases to 94.9%. After removing all modules, Mask R-CNN (ResNet) has the weakest showing an accuracy of 94.3% and an F1 value of 93.4%.

D. Experimental Discussion

By introducing modules such as ResNeXt, CBAM and Boundary, the improved Mask R-CNN model has significantly enhanced its capabilities in feature extraction, attention mechanism and precise positioning of target boundaries. The ResNeXt module improves the recognition ability of the model through more efficient feature learning. The CBAM module effectively focuses on important areas in the image and enhances the performance of the model in complex backgrounds. The Boundary loss function ensures higher segmentation accuracy by improving boundary detection accuracy. The improved Mask R-CNN has a slow frame rate and convergence speed, mainly due to the complex model architecture and fine-grained image segmentation tasks that lead to large computational workload.

In real-time applications, improving accuracy and reducing frame rate require a trade-off between computational complexity and detection accuracy. The improved Mask R-CNN improves recognition accuracy through modules such as ResNeXt, CBAM, and Boundary loss, but the additional computational overhead leads to a low frame rate, which makes it difficult to meet high real-time requirements. To balance the two, the following optimization strategies will be adopted:

(1) Combine model pruning, quantization, and distillation techniques to reduce computational redundancy and speed up inference;

(2) Use more efficient inference frameworks, such as TensorRT or ONNX Runtime, to accelerate model inference;

(3) Adjust the model for different application scenarios. Under high real-time requirements, lightweight networks such as MobileNet or YOLO series can be used for fast detection, while improved Mask R-CNN can be used for fine recognition in scenarios with high precision requirements.

The experimental outcomes of this study offer new ideas for safety monitoring in the field of hazardous operation action recognition, especially in the power industry. The study improves the robustness and accuracy of the system in practical applications and also provides a reference model optimisation path for intelligent monitoring systems in other fields. This study verified the advantages of modular design in deep learning models and found that the improvement of model performance does not only rely on a single network structure, but on the coordinated optimization of multiple modules. The method has good promotion value, especially in the fields of industrial safety, medical diagnosis and intelligent monitoring.

The dataset in this paper has certain limitations. In order to improve the versatility of the model, more diverse data enhancement strategies will be adopted in the future, such as random cropping, rotation, color jitter, and lighting changes, to simulate power operation scenes under different environmental conditions and improve the generalization ability of the model. Introduce more real-world data, covering different weather, lighting, shooting angles, and equipment types, to reduce the model's dependence on specific scenes. Transfer learning is also an effective method. Models pre-trained on larger general datasets such as industrial action recognition datasets can be used and fine-tuned on power operation data to enhance adaptability to new scenarios.

This study has made significant progress, but there are still some limitations:

(1) The dataset used in this study is mainly concentrated in the power sector, and the data scale is relatively small, lacking effective verification of the model's generalization ability. In the future, the universality of the model can be further verified through diversified datasets, including other industries and more types of dangerous operations.

(2) The improved Mask R-CNN performs well in recognition accuracy, but its frame rate is low, which limits its performance in real-time applications. In the future, we will further use model quantization FP16 and pruning technology to reduce inference latency, and use multi-threading or GPU parallel computing to optimize the inference process, while combining efficient inference engines such as TensorRT or ONNX Runtime to accelerate computing. Finally, we will use distillation learning to train smaller but more efficient models, while ensuring accuracy and improving frame rate.

(3) There are many types of dangerous operations in power marketing and energy metering. This paper only discusses some of them. In the future, more dangerous operations can be collected to enrich the data set.

4. Conclusions

This paper adopts an improved Mask R-CNN image segmentation model for the recognition of dangerous operation actions at the power marketing and energy metering site. The study introduces the GAN model for data expansion, uses the ResNeXt module to replace the traditional ResNet for feature extraction, embeds the CBAM module to reduce background interference, and combines the Boundary loss function to optimize the boundary accuracy, significantly improving the recognition accuracy and segmentation accuracy. The outcomes show that the enhanced mask R-CNN performs

significantly superior to the traditional CNN model in real power operation site images, and is able to effectively identify and accurately segment dangerous operation behaviours to ensure the safety of the operators.

This paper has made some achievements, but there are also some shortcomings. The data scale is relatively small, and the generalization ability of the model is not effectively verified. In addition, the types of dangerous operations at the power marketing and energy metering site are not considered comprehensively, and the calculation efficiency is not particularly ideal. In the future, more types of dangerous operations will be collected to expand the data set, and the model will be quantized using FP16 and pruning technology to optimize the network structure to improve convergence performance and calculation efficiency.

Acknowledgment

None

Consent to Publish

The manuscript has neither been previously published nor is under consideration by any other journal. The authors have all approved the content of the paper.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author, upon reasonable request.

Funding

None

Author Contribution

Lei Wei: Developed and planned the study, performed experiments, and interpreted results. Edited and refined the manuscript with a focus on critical intellectual contributions.

Liang Wang, Feihong Yin: Participated in collecting, assessing, and interpreting the date. Made significant contributions to date interpretation and manuscript preparation.

Junchen Guo: Provided substantial intellectual input during the drafting and revision of the manuscript.

Conflicts of Interest

The authors declare that they have no financial conflicts of interest.

References

[1] Z.Y. Chen, A.M. Amani, X.H. Yu, M. Jalili. Control and optimisation of power grids using smart meter data: A review. Sensors, 2023, 23(4), 2118-2143. DOI: 10.3390/s23042118

[2] T. Knayer, N. Kryvinska. An analysis of smart meter technologies for efficient energy management in households and organizations. Energy Reports, 2022, 8(1), 4022-4040. DOI: 10.1016/j.egyr.2022.03.041

[3] M.A. Saeed, A.A. Eladl, B.N. Alhasnawi, S. Motahhir, A. Nayyar, et al. Energy management system in smart buildings based coalition game theory with fog platform and smart meter infrastructure. Scientific Reports, 2023, 13(1), 2023-2039. DOI: 10.1038/s41598-023-29209-4

[4] Y. Ke, H.T. Chen, Z.Y. Liu, Z.Y. Yang, L. Song. Research on three-dimensional perception and protection technology for power construction safety operations. International Journal of Wireless and Mobile Computing, 2024, 27(2), 133-140. DOI: 10.1504/IJWMC.2024.140270

[5] P.S. Duan, J.L. Zhou. Cascading vulnerability analysis of unsafe behaviors of construction workers from the perspective of network modeling. Engineering, Construction and Architectural Management, 2023, 30(3), 1037-1060. DOI: 10.1108/ECAM-06-2021-0475

[6] G.B. Wang, M.Y. Liu, D.P. Cao, D. Tan. Identifying high-frequency–low-severity construction safety risks: an empirical study based on official supervision reports in Shanghai. Engineering, Construction and Architectural Management, 2022, 29(2), 940-960. DOI: 10.1108/ECAM-07-2020-0581

[7] W.T. Xin, R.Y. Liu, Y. Liu, Y. Chen, W.X. Yu, et al. Transformer for skeleton-based action recognition: A review of recent advances. Neurocomputing, 2023, 537(1), 164-186. DOI: 10.1016/j.neucom.2023.03.001

[8] A.M. Helmi, M.A.A. Al-qaness, A. Dahou, M. Abd Elaziz. Human activity recognition using marine predators algorithm with deep learning. Future Generation Computer Systems, 2023, 142(1), 340-350. DOI: 10.1016/j.future.2023.01.006

[9] W.J. Yang, J.L. Zhang, J.J. Cai, Z.X. Xu. HybridNet: Integrating GCN and CNN for skeleton-based action recognition. Applied Intelligence, 2023, 53(1), 574-585. DOI: 10.1007/s10489-022-03436-0

[10] J. Lee, S. Lee. Construction site safety management: a computer vision and deep learning approach. Sensors, 2023, 23(2), 944-965. DOI: 10.3390/s23020944

[11] M.M. Alateeq, F.R. P.P., M.A.S. Ali. Construction site hazards identification using deep learning and computer vision. Sustainability, 2023, 15(3), 2358-2375. DOI: 10.3390/su15032358

[12] M. Park, D.Q. Tran, J. Bak, S. Park. Small and overlapping worker detection at construction sites. Automation in Construction, 2023, 151(10), 1-14. DOI: 10.1016/j.autcon.2023.104856

[13] I.U. Khan, S. Afzal, J.W. Lee. Human activity recognition via hybrid deep learning based model. Sensors, 2022, 22(1), 323-338. DOI: 10.3390/s22010323

[14] N.R. Malik, S.A.R. Abu-Bakar, U.U. Sheikh, A. Channa, N. Popescu. Cascading pose features with CNN-LSTM for multiview human action recognition. Signals, 2023, 4(1), 40-55. DOI: 10.3390/signals4010002

[15] I. Shah, H. Iftikhar, S. Ali. Modeling and forecasting electricity demand and prices: A comparison of alternative approaches. Journal of Mathematics, 2022, 2022(1), 3581037. DOI: 10.1155/2022/3581037

[16] H. Iftikhar, J. Zywiołek, J.L. Lopez-Gonzales, O. Albalawi. Electricity consumption forecasting using a

novel homogeneous and heterogeneous ensemble learning. Frontiers in Energy Research, 2024, 12, 1442502. DOI: 10.3389/fenrg.2024.1442502
[17] S.M. Gonzales, H. Iftikhar, J.L. Lopez-Gonzales. Analysis and forecasting of electricity prices using an improved time series ensemble approach: An application to the Peruvian electricity market. AIMS Mathematics, 2024, 9(8), 21952-21971. DOI: 10.3934/math.20241067
[18] H. Iftikhar, S.M. Gonzales, J. Zywiołek, J.L. Lopez-Gonzales. Electricity demand forecasting using a novel time series ensemble technique. IEEE Access, 2024, 12, 88963-88975. DOI: 10.1109/ACCESS.2024.3419551
[19] J.Q. Li, G.Y. Zhou, D.F. Li, M.Y. Zhang, X.F. Zhao. Recognizing workers' construction activities on a reinforcement processing area through the position relationship of objects detected by faster R-CNN. Engineering, Construction and Architectural Management, 2023, 30(4), 1657-1678. DOI: 10.1108/ECAM-04-2021-0312
[20] X. Li, T.X. Hao, F. Li, L.Z. Zhao, Z.H. Wang. Faster r-cnn-lstm construction site unsafe behavior recognition model. Applied Sciences, 2023, 13(19), 1-16. DOI: 10.3390/app131910700
[21] Gaurav, S. Bhardwaj, R. Agarwal. An efficient speaker identification framework based on Mask R-CNN classifier parameter optimized using hosted cuckoo optimization (HCO). Journal of Ambient Intelligence and Humanized Computing, 2023, 14(10), 13613-13625. DOI: 10.1007/s12652-022-03828-7
[22] R. Rijayanti, M. Hwang, K. Jin. Detection of Anomalous Behavior of Manufacturing Workers Using Deep Learning-Based Recognition of Human–Object Interaction. Applied Sciences, 2023, 13(15), 8584-8597. DOI: 10.3390/app13158584
[23] R. Dhivy, S.V. Kogilavani. A Mask-Based Recurrent Convolutional Neural Network for Moving Object Detection in Surveillance Videos. Journal of the Balkan Tribological Association, 2024, 30(3), 321-334.
[24] U. Gawande, K. Hajari, Y. Golhar. Real-time deep learning approach for pedestrian detection and suspicious activity recognition. Procedia Computer Science, 2023, 218(1), 2438-2447. DOI: 10.1016/j.procs.2023.01.219
[25] G.F. Li, B.A. Li, D. Jiang, B. Tao, J.T. Yun. An improved mask R-CNN example segmentation algorithm based on RGB-D. International Journal of Wireless and Mobile Computing, 2024, 26(3), 302-309. DOI: 10.1504/IJWMC.2024.137861
[26] Sumit, S. Bisht, S. Joshi, U. Rana. Comprehensive Review of R-CNN and its Variant Architectures. International Research Journal on Advanced Engineering Hub (IRJAEH), 2024, 2(04), 959-966. DOI: 10.47392/IRJAEH.2024.0134
[27] B. Goyal, A. Gupta, A. Dogra, D. Koundal. An adaptive bitonic filtering based edge fusion algorithm for Gaussian denoising. International Journal of Cognitive Computing in Engineering, 2022, 3(1), 90-97. DOI: 10.1016/j.ijcce.2022.03.001
[28] G.S.B. Nitin. A hybrid image denoising method based on discrete wavelet transformation with pre-gaussian filtering. Indian Journal of Science and Technology, 2022, 15(43), 2317-2324. DOI: 10.17485/IJST/v15i43.1570
[29] K. Silva, B. Can, R. Sarwar, F. Blain, R. Mitkov. Text data augmentation using generative adversarial networks–a systematic review. Journal of Computational and Applied Linguistics, 2023, 1(1), 6-38. DOI: 10.33919/JCAL.23.1.1
[30] A.G. Dieste, F. Arguello, D.B. Heras. Resbagan: A residual balancing gan with data augmentation for forest mapping. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16(1), 6428-6447. DOI: 10.1109/JSTARS.2023.3281892
[31] A. Bousmina, M. Selmi, M.A. Ben Rhaiem, I.R. Farah. A hybrid approach based on gan and cnn-lstm for aerial activity recognition. Remote Sensing, 2023, 15(14), 3626-3645. DOI: 10.3390/rs15143626
[32] B.H. Zou, H.Z. Yan, F.Q. Wang, Y.C. Zhang, X.D. Zeng. Research on signal modulation classification under low SNR based on ResNext network. Electronics, 2022, 11(17), 2662-2672. DOI: 10.3390/electronics11172662
[33] A. Haryono, G. Jati, W. Jatmiko. Oriented object detection in satellite images using convolutional neural network based on ResNeXt. ETRI Journal, 2024, 46(2), 307-322. DOI: 10.4218/etrij.2022-0446
[34] S. Mekruksavanich, A. Jitpattanakul. Deep Residual Network with a CBAM Mechanism for the Recognition of Symmetric and Asymmetric Human Activity Using Wearable Sensors. Symmetry, 2024, 16(5), 554-579. DOI: 10.3390/sym16050554
[35] Y. Wang, X.Q. Chen, J.Q. Li, Z.X. Lu. Convolutional Block Attention Module–Multimodal Feature-Fusion Action Recognition: Enabling Miner Unsafe Action Recognition. Sensors, 2024, 24(14), 4557-4574. DOI: 10.3390/s24144557
[36] M. Munsif, S.U. Khan, N. Khan, S.W. Baik. Attention-based deep learning framework for action recognition in a dark environment. Human-centric Computing and Information Sciences, 2024, 14(1), 1-22. DOI: 10.22967/HCIS.2024.14.004
[37] E. Hassan, M.Y. Shams, N.A. Hikal, S. Elmougy. The effect of choosing optimizer algorithms to improve computer vision tasks: a comparative study. Multimedia Tools and Applications, 2023, 82(11), 16591-16633. DOI: 10.1007/s11042-022-13820-0
[38] N.C. Xiao, X.Y. Hu, X. Liu, K.C. Toh. Adam-family methods for nonsmooth optimization with convergence guarantees. Journal of Machine Learning Research, 2024, 25(48), 1-53. DOI: 10.48550/arXiv.2305.03938
[39] M. Reyad, A.M. Sarhan, M. Arafa. A modified Adam algorithm for deep neural network optimization. Neural Computing and Applications, 2023, 35(23), 17095-17112. DOI: 10.1007/s00521-023-08568-z