

# Research on Wear Recognition of Electric Worker's Helmet Based on Neural Network

Yuwei Sun\*, Jing Fu, Qingshan Ma, Xinchun Yu, Guoxiang Fan

Nantong Power Supply Branch of State Grid Jiangsu Electric Power Co., Ltd.,  
Nantong, Jiangsu, 226006, China

\*Corresponding author's e-mail: yuxinchuntpower@163.com

**Abstract.** Head injury is an important cause of electric construction accidents. Wearing a safety helmet is an effective measure to prevent electric construction accidents, and in the work, unsafe behavior of workers without wearing a helmet often occurs. This paper proposes a neural network-based automatic cap wear automatic identification and detection technology to identify the behavior of unwearing helmets in power construction. In view of the shortcomings of the existing helmet detection scheme in the complex environment, such as low efficiency and low recognition accuracy, this paper selects the region of the neural network (Fast-er-R-CNN) to identify the wearing of the helmet during the power operation. The overall scheme uses the target detection scheme based on the foreground detection technology to detect the target individual in the obtained video recording of the power operation. The Kalman filter algorithm is used to track the target in the video and solve the occlusion problem in the video image. Through the actual detection of power construction monitoring video, the availability and efficiency of the designed algorithm are verified. The target recognition accuracy can reach 85.7% and the recall rate is 87.5%.

## 1. Introduction

With the continuous development of deep learning technology, video image processing technology has been widely used in various fields of social life [1]. For the power industry, security is not unfamiliar to everyone. Only by achieving security can the company ensure stable production, and the losses caused by security incidents are enormous [2]. The helmet has a certain protective effect on the head during power work, which requires the electrician to wear a helmet during the construction process. However, in recent years, electric safety accidents caused by electricians who do not wear safety helmets in violation of electric safety regulations have occurred from time to time. In order to prevent the occurrence of electric safety accidents and protect the personal safety of electric power workers, it is becoming more and more important for the electric power industry to develop a system that can automatically identify abnormal situations in which electricians do not wear helmets during operation.

In recent years, some experts and scholars have studied the identification technology of helmets. Wei Hu uses the BP neural network and wavelet transform method to identify the application of the helmet [3]. This method relies heavily on the positional relationship between the head and the face when extracting features, and the detection efficiency is low in complex environments. Yunbo Liu conducts the identification of the helmet by the color of the target detection pixel [4]. Lin Bi recognizes the completed helmet through the convolutional neural network [5], but the image features can only identify the shape and color of the helmet. The wearing condition of the helmet is judged, and the recognition accuracy of the method to the target object is not high. The region-convolutional



neural network (R-CNN) [6] proposed in 2014 combines deep learning with target detection to achieve high-precision target classification by layer-by-layer extraction of image features. However, due to the long time and space complexity of the method, in 2016, based on R-CNN), the step-by-step optimization such as Fast R-CNN [7] and Faster R-CNN) [8] greatly reduced the complexity of time and space. This paper analyzes the application requirements of the power industry and selects Faster R-CNN as the basic network model for the helmet identification and tracking detection algorithm. The overall design of this paper is mainly composed of three parts: target detection based on foreground detection, target individual identification and target tracking. While performing foreground processing on the video image of the electric power operator, the trained target detector is applied to the image subjected to the foreground detection processing and the target individual is identified by using the video image training target detector. Finally, the experimental results verify the availability and efficiency of the designed solution by conducting experiments on surveillance videos from the power industry for safe operations.

Target detection based on foreground detection Video-based target detection scheme Compared with the target detection task of static image, the appearance, shape, scale and other attributes of the target will change with the movement of the target. The main difficulty of the video target detection task is how to maintain the consistency of the target in the time sequence during the detection process so that the target is not lost in the frame. Specifically, there are five main challenges in this aspect: Motion blur, virtual focus, occlusion, appearance change, scale change. According to the surveillance video data, the main challenges encountered this time are occlusion and scale changes. As shown in Figure 1, the target's appearance changes during the movement from the far side of the lens to the vicinity.



Figure 1. The scale of the target changes

In order to solve the problem of target recognition and target tracking caused by dimensional changes and to extract the target from the image content with complex background in the surveillance video, we chose to use foreground segmentation technology. The most widely used model for foreground detection is the Gaussian Mixture Model (GMM). Therefore, we use a mixed Gaussian model for foreground segmentation of power-work video images. The earliest application of the mixed Gaussian model in computer vision is the foreground detection in the field of video surveillance. This model is not only stable but also self-learning. Before performing foreground detection, the background image of the power working environment is trained, and a mixed Gaussian model is used to simulate each background in the image. Then, in the test phase, the new image is mixed Gaussian model matching, and if the pixel value of the image can be matched, it is determined by the computer as the background image, otherwise it is recognized as the foreground image. Since the GMM model is continuously updated and learned throughout the process, it is robust to dynamic backgrounds.

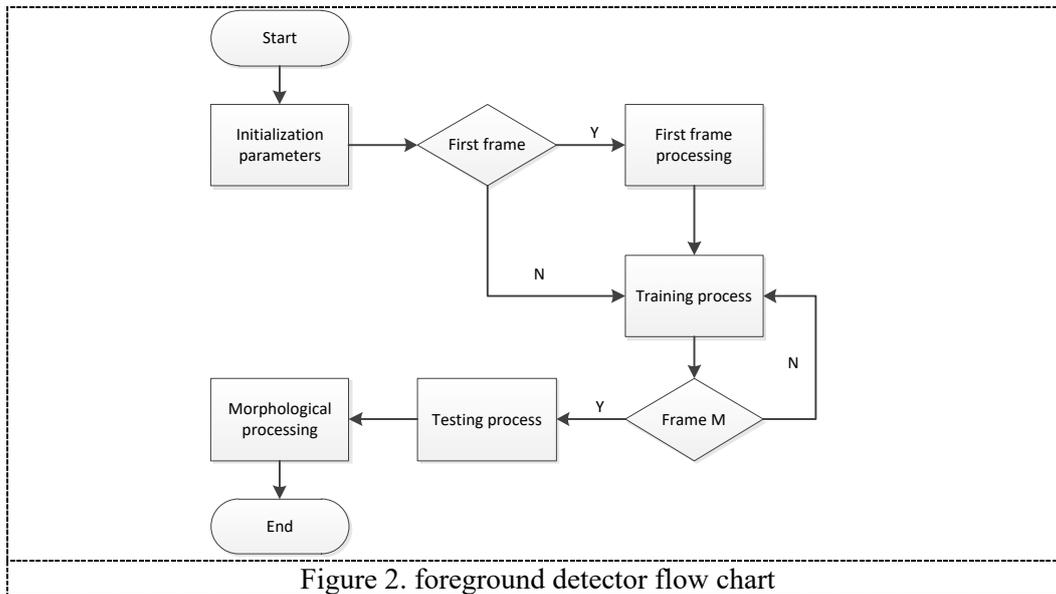


Figure 2. foreground detector flow chart

But the foreground segmentation process is not perfect and usually includes unwanted noise, we use the morphological opening function in morphology to eliminate noise and fill the gaps in the detected object. The process firstly etches the edge of the image by etching the image, thereby kicking off the burr of the edge of the target, and then expanding the edge of the image by an expansion algorithm to complement the missing portion of the edge of the target. Using the same number of corrosion and expansions, the target surface can be smoothed. The specific design flow of the foreground detector is shown in Figure 2.

## 2. Kalman filter algorithm

In order to solve the occlusion problem, the target tracking algorithm we adopted is the Kalman filter algorithm. Kalman filtering can be used in dynamic systems with uncertain information to make an informed prediction of the next step of the system. Even with various disturbances, Kalman filtering can indicate the actual situation. Here, we apply Kalman filtering to predict the position of the individual in the power operation monitoring video, predict the position of the next moment by the current position of the subject, and then correct the prediction result by the actual position of the next moment of the object to Get the optimal predicted position. With this method, when the actual position is not available due to occlusion, this method can be used to predict its position so as not to lose the tracking target. The main flow of the Kalman filter algorithm can be divided into the following five steps:

1. Predicting the position of the target at time  $k$  from the target position at time  $k-1$ ;

$$\hat{x}_{k/k-1} = \phi_{k,k-1}^* * x_{k-1} + B_{k-1} u_{k-1} + J_{k-1} Z_{k-1} \quad (1)$$

2. Predict new errors by existing error covariance and process noise  $Q$ ;
3. Calculating Kalman gain;

$$H_k = \hat{P}_{k/k-1} * C_k^T [C_k * \hat{P}_{k/k-1} * C_k^T + R_k]^{-1} \quad (2)$$

4. Perform calibration update, perform filter estimation, and obtain the optimal position at time  $k$ ;
5. Filtering the mean square error update operation for the next iteration of estimating the optimal position at time  $k+1$ .

$$P_k = [I - H_k * C_k] * \hat{P}_{k/k-1}$$

(3)

### 3. Faster R-CNN Network model training

After detecting and identifying the target individual in the power operation video image, the Faster R-CNN network model is used for training to classify the individual wearing the helmet in the image. The object detection scheme based on Faster R-CNN network is a scheme in which the R-CNN domain is better in object detection and faster in training. The specific training process of the Faster R-CNN network model is shown in Figure 3. The power operation video image is taken as input and passed to the convolutional neural network, which returns the feature map of the image, and the RPN is applied to the feature map. Return the proposed object and its score, then apply the ROI pool layer on these proposed objects to reduce all proposals to the same size. Finally, the proposal is passed to the fully connected layer, which has a soft max layer and a linear regression layer to classify and output the borders of the objects. The convolutional network model in Faster R-CNN is a pre-trained CNN model, where we use pretrained ResNet-50 for feature extraction in convolutional networks. Next, we use the Selective Search [9] for candidate region selection, which is currently the best-performing region candidate algorithm.

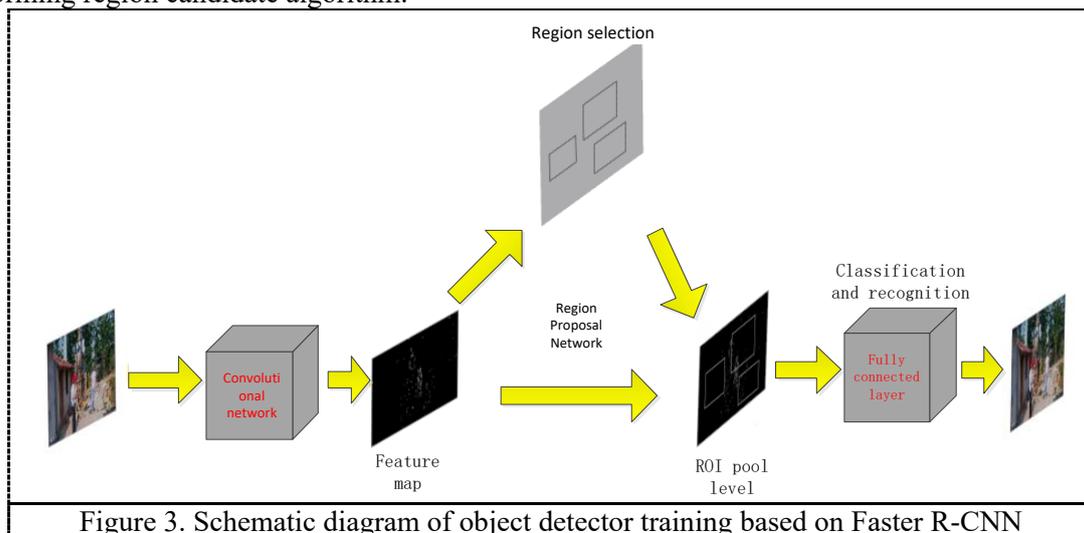


Figure 3. Schematic diagram of object detector training based on Faster R-CNN

### 4. Faster R-CNN based helmet identification

The helmet identification scheme based on Faster R-CNN is mainly divided into three components: foreground detection, target detection and target tracking. The specific relationship is shown in Figure 4. The foreground detector is used to separate the object to be identified from the complex background image and improve the accuracy of target recognition. The purpose of the target detector is to identify objects that wear helmets and no helmets. The target tracking technique is to not lose the identified target information in the video [10]. While performing foreground processing on the power operation video image, the image training target detector is used to apply the trained target detector to the image subjected to the foreground detection processing and target recognition thereof.

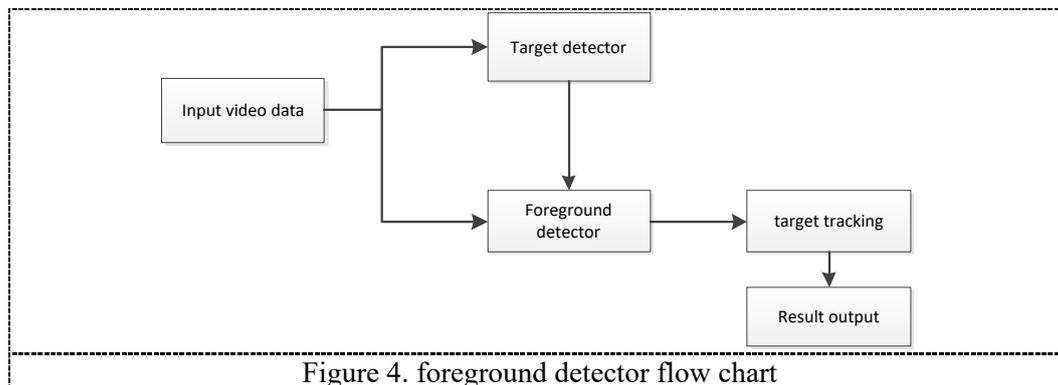


Figure 4. foreground detector flow chart

First, a Gaussian mixture model is designed to detect the target in the foreground detector. The Foreground Detector function was used during the import of the video and the installation of the foreground detector, and the three Gaussian modes were initialized in the hybrid model with the first fifty frames of the video. Use the morphological opening to eliminate noise and fill the gaps in the detected object. The function `imopen` is used. Use the Blob Analysis function to find the bounding box of each Unicom component that corresponds to the moving object. This function further filters the detected foreground by rejecting spots containing less than 50 pixels. Secondly, the object detector based on Faster R-CNN is trained, the image data set of the character is decompressed, the ROI tag is loaded and inserted, and the data is randomly divided into the training set and the test set by the `randperm` function, and the training options are configured:

```

options = trainingOptions('sgdm', ...
    'MiniBatchSize', 1, ...
    'InitialLearnRate', 1e-4, ...
    'LearnRateSchedule', 'piecewise', ...
    'LearnRateDropFactor', 0.1, ...
    'LearnRateDropPeriod', 100, ...
    'MaxEpochs', 20, ...
    'CheckpointPath', tempdir, ...
    'Verbose', true);
  
```

The Trainer RNN-based object detector is trained using `trainFasterRCNNObjectDetector` `hanshu1`, where the feature selection network selects ResNet-50. Next, the video processed in the first step is imported, framed, and the target is detected by the trained target detector, and the multi-target is tracked by the Kalman filter algorithm. Use the `detectObjects` function to import the trained detector to detect the target, and use the `PredictNewLocationsOfTracks` function to predict the centroid of each track in the current frame using the Kalman filter and update its bounding box accordingly. Use the distance method of the `SystemObject` function to calculate the cost of each test for each track. This cost takes the Euclidean distance between the predicted centroid of the orbit and the detected centroid into account, it also includes the confidence of the prediction, which is maintained by the Kalman filter, and the result is stored in an  $M \times N$  matrix, where  $M$  is the orbit Number,  $N$  is the number of detections. Then use the new detection to correct the estimate of the object's position and then replace the predicted bounding box with the detected bounding box.

Finally, the trained target detector is applied to the image after foreground detection processing, and the target person who wears the helmet and without wearing the helmet is targeted, and the target tracking technology is used to perform target tracking on the identified individual object.

## 5. Experimental results

### 5.1. Training the RPN layer in the Faster R-CNN model

Training on a single-core GPU, the training process is iterated 20 times, cycled 380 times, the accuracy rate from the initial 63% through continuous cycling and iteration can finally reach 100%,

the learning rate is 1.0000e-04, The loss is 0.495 and the RMSE value is 0.71. The specific training process is shown in Figure 5.

```

Training a Faster R-CNN Object Detector for the following object classes:

* manwithHelmet

Step 1 of 4: Training a Region Proposal Network (RPN).
Training on single GPU.
=====
| Epoch | Iteration | Time Elapsed | Mini-batch | Mini-batch | Mini-batch | Base Learning |
|       |          | (hh:mm:ss)  | Loss       | Accuracy   | RMSE       | Rate         |
|=====|=====|=====|=====|=====|=====|=====|
| 1     | 1       | 00:00:01    | 1.3587     | 63.28%    | 0.95       | 1.0000e-04  |
| 3     | 50      | 00:01:01    | 0.0695     | 100.00%   | 1.05       | 1.0000e-04  |
| 6     | 100     | 00:02:17    | 0.2618     | 100.00%   | 0.47       | 1.0000e-04  |
| 8     | 150     | 00:03:24    | 0.1187     | 100.00%   | 0.40       | 1.0000e-04  |
| 11    | 200     | 00:04:58    | 0.1087     | 100.00%   | 1.24       | 1.0000e-04  |
| 14    | 250     | 00:06:17    | 0.3507     | 100.00%   | 0.59       | 1.0000e-04  |
| 16    | 300     | 00:07:10    | 0.0800     | 100.00%   | 0.32       | 1.0000e-04  |
| 19    | 350     | 00:08:29    | 0.0708     | 100.00%   | 0.32       | 1.0000e-04  |
| 20    | 380     | 00:09:01    | 0.0495     | 100.00%   | 0.71       | 1.0000e-04  |
|=====|=====|=====|=====|=====|=====|

```

Figure 5. Faster R-CNNRPN layer training in the network model

### 5.2. Target extraction based on foreground segmentation technology

Next, the target of extraction step based on the fore-ground segmentation technique is verified. Figure 6(a) is a frame extracted from the surveillance video image captured by the power industry during the construction process. The target is extracted from the complex video image background. The result is shown in Figure 6(b). The final result after noise elimination is shown in Figure 6(c).

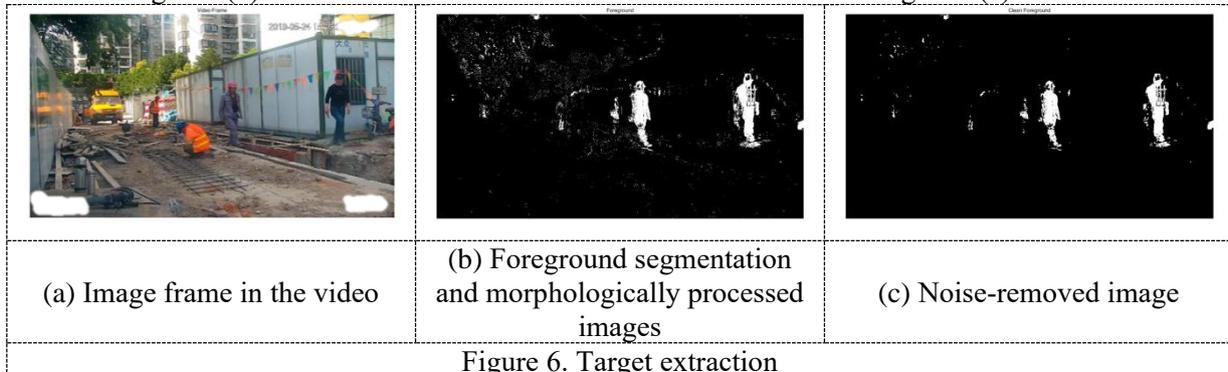


Figure 6. Target extraction

### 5.3. Target recognition and tracking effect based on Faster R-CNN

Then use the trained target-based detector to perform target recognition and tracking processing on the video images processed by the foreground segmentation technology. After experimental tests, the accuracy rate can reach 85.7%, and the recall rate is 87.5%. The detection effect of the single-frame image in the video is as shown in FIG. 7, and the green frame and the blue frame respectively indicate the case where the helmet is properly worn as required and the helmet is not worn as required.



Figure 7. Target detection results for single-frame images in video

Finally, using the target tracking technology based on Kalman filter algorithm to locate and track the identified targets. As shown in Figure 8, it selected three different time states in the surveillance video for identification and tracking. 8(a), 8(b) and 8(c) show the results of the construction personnel being identified and tracked in different time states 1, time state 2 and time state 3, respectively.



(a) Tracking result in state 1

(b) Tracking result in state 2

(c) Tracking result in state 3

Figure 8. Target tracking effect in video

## 6. Conclusion

Based on the deep learning technology, this paper realizes the automatic identification and detection of the helmet wearing situation in the power industry during safe operation. Through the analysis of the application requirements of the power industry, the Faster R-CNN is selected as the basic network model of the helmet identification algorithm for training. Then, combined with the target extraction method based on foreground segmentation technology and Kalman filter algorithm, the automatic identification detection and tracking of the helmet wearing situation in the power construction video is realized. Finally, the availability and efficiency of the designed algorithm are verified by experiments on surveillance video from the power industry for safe operation. The research in this paper explores an effective and feasible path for the automatic recognition and tracking of deep learning technology in power security operations.

## References

- [1] Zou Ruisheng, Xie Yingwen, Wei Chaobing, Li Xiaobin. Research on the application of face recognition technology in the field of power safety construction [J]. Digital communication world, 2017 (10): 64.
- [2] Huang Wenqi, Zhang Fuzheng, Li Peng, Ming Zhe, Xu Aidong, Chen Huajun, Yang hang. Insulator identification exploration and application based on fast r-cnn [J]. China Southern Power Grid technology, 2018,12 (09): 22-27.
- [3] Hu Tian, Wang Xingang. Analysis and design of helmet recognition system based on wavelet transform and neural network. Software guide, 2006 (23): 37-39.
- [4] Liu Yunbo, Huang Hua. Research on monitoring technology of wearing safety helmet on construction site [ J ] . Electronic technology, 2015, 28 (4): 69-72.
- [5] Bi Lin, Xie Wei, Cui Jun. research on recognition of safety helmet wearing of miners based on convolutional neural network [J]. Gold science and technology, 2017,25 (04): 73-80.

- [6] Girshick, R., J. Donahue, T. Darrell, J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. [C] IEEE Conference on Computer Vision and Pattern Recognition. Columbus, OH, 2014: 772–779.
- [7] Girshick, R. "Fast R-CNN." Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile, Dec. 2015, pp. 1440-1448.
- [8] Ren, S., K. He, R. Gershick, and J. Sun. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 39, Issue 6, June 2017, pp. 1137-1149.
- [9] Uijlings, J. R. R., K. E. A. van de Sande, T. Gevers, and A. W. M. Smeulders. "Selective Search for Object Recognition." International Journal of Computer Vision. Vol. 104, Number 2, Sept. 2013, pp. 154-171.
- [10] Zitnick, C. L., and P. Dollar. "Edge Boxes: Locating Object Proposals from Edges." European Conference on Computer Vision. Zurich, Switzerland, Sept. 2014, pp. 391-405.