PERSONNEL TRACKING OF POWER CONSTRUCTION SITE METHOD BASED ON F-TRACKTOR++

Zihao LIU*, Jining ZHAO, Haifeng LIU, Rong MENG, Zhilong ZHAO

With the increasing dispersion of power construction sites, increasing intelligence of construction equipment, and upgrading of management personnel's demand for recording construction process, intelligent methods are needed to monitor and record the trajectories of construction personnel. To address this, the Tracktor++ algorithm is improved to be the Feature-enhanced Tracktor++ (F-Tracktor++) algorithm for power construction. First, target recognition is optimized by using Resnet-50 and feature pyramid module, and the SimAM attention mechanism is introduced. Then the motion estimation part is replaced with Kalman filter for processing. Finally, the network structure of the re-identification part is improved by taking ResNeXt as the backbone network and a batch normalization neck network as the detection head. After validation, compared with the original Tracktor++ net-work, the improved Tracktor++ algorithm can achieve a 3.2% reduction in the overall false positive value, a 12.3% reduction in identity switches, and a 90% increase in successful tracking ratio. The accuracy of tracking trajectory can be ensured through improving the accuracy of personnel identification and recognition greatly. The improved Tracktor++ algorithm has good performance in personnel positioning and tracking at power construction sites, achieving automatic positioning and accurate tracking of personnel.

Keywords: Transmission engine, deep learning, convolutional neural network, target tracking

1. Introduction

The increasing demand for electricity in daily life and industry has posed great challenges to the reliability and safety of the distribution system. According to the specifications of power operation, construction personnel should construct in the specified range [1-2]. If construction personnel mistakenly enter an unsafe space during construction, safety accidents may occur, endangering the safety of personnel and equipment operation. Therefore, it is necessary for safety supervisors or construction managers to monitor the location of construction personnel in real time. This work is mainly completed through on-site supervision by supervisors or managers. With the development of power construction, construction sites have shown the characteristics of large quantity, wide scope, and complex and different environment. At the same time, the demnadfor controlling construction costs and recording construction processes is also increasing [3]. It is necessary to adopt non-

* State Grid Hebei Electric Power Co., Ltd. Ultra High Voltage Branch, Shijiazhuang, Hebei, China, E-mail: hvauyyq@163.com

manual, intelligent methods to monitor and record the activities of construction workers.

There are often two common non-artificial positioning tracking methods: hardware positioning tracking and visual positioning tracking. *Madavarapu* used the Global Positioning System (GPS) technology to continuously collect the location of personnel for tracking [4]. Sarmento et al. used the Ultra Wide Band (UWB) technology to calculate the distance and position of personnel wearing tag devices by measuring the time difference of signal transmission between base stations and tag devices, achieving the ability of robots to follow humans. Yang et al. collected data using millimeter wave radar and generated human motion trajectories based on the maximum energy using beam forming algorithm [6]. The above methods partially meet the requirements of on-site positioning and tracking, but with high cost. In addition, other technologies such as GPSand UWB all use wireless signals for positioning, which are easily affected by electromagnetic interference and obstacles [7]. The method of locating and tracking construction personnel through hardware implementation has certain drawbacks.

Visual positioning and tracking methods use cameras placed at the construction site to capture videos. The positioning and tracking algorithms can provide positioning and tracking functions for construction personnel. Traditional target tracking algorithms are generally composed of manually designed features and shallow appearance models. The aim is to use simple and effective visual features and shallow matching or classification models to design fast and robust tracking algorithms. For example, a histogram of silhouettes and optical flow is created to extract the shape and motion information of human bodies in images, and an improved support vector machine algorithm is used for feature fusion [1]. In various complex scenarios, due to the interference such as lighting changes and personnel movement, the same targets have different patterns in different images [8]. The tracking accuracy of traditional algorithms are greatly limited.

Deep learning methods have achieved optimal feature modeling ability in object detection and tracking. Some scholars have used deep learning methods to improve the intelligence of power safety monitoring. Miao et al. combined shuffle attention and Focal IoU loss function to improve the performance of YOLOv8 [8]. The method can achieve the detection and classification of personnel behavior. However, it cannot distinguish different individuals, as well as track and record personnel trajectories. Further research is needed on personnel detection and tracking methods.

Based on the real situation of power construction site, the Tracktor++ algorithm [9] is improved to the Feature-enhanced Tracktor++ algorithm (F-Tracktor++) for personnel positioning and tracking on power construction sites. The target detection part of the Tracktor++ algorithm is first optimized by using Resnet-50 and Feature Pyramid Networks (FPN) [10], and the SimAM attention

module [11] is introduced. Then the motion estimation part is improved by using a Kalman filter. Finally, the re-identification (ReID) network structure of the Tracktor++ algorithm is modified to introduce the ResNeXt network as the backbone network and the Batch Normalize Neck (BNNeck) as the detection head, further improving the stability of recognition.

The dataset obtained from the electric power construction site is used for training and experiments. The experimental results showed that compared with the original Tracktor++network, the False Positive (FP) value of the F-Tracktor++ algorithm is reduced by 3.2%; the Identity Switches (IDs) value is reduced by 12.3%, and the Mostly Tracked (MT) ratio is increased to 90%.

2. Material and Research Method

2.1 Tracktor++ Algorithm

Commonly used target tracking algorithms such as SORT [12], DeepSort [13] and StrongSort [14] are plagued by high complexity and computation. To address these limitations, Bergmann et al. [9] proposed the Tracktor algorithm for multi-object tracking (MOT), which integrates detection and tracking into a unified network framework while embedding detection and identity features. This design enables multi-task learning through shared components, thereby improving the performance of the tracker. The tracking process of the Tracktor algorithm is mainly divided into two parts, as shown in Fig. 1.

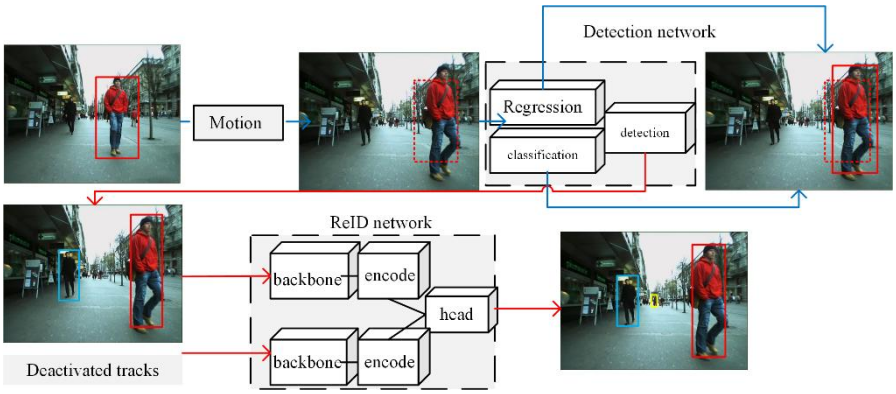

Fig. 1 Steps of Tracktor++ for multi-object tracking

The first part is targeting box regression, as indicated by the blue arrow in Fig. 1. The bounding box regression is performed on the target of frame t-1 to obtain the new position of the target in frame t. Using the Faster-RCNN [10] as the target detection network, region of interest pooling is conducted with the target boxes in frame t-1, and then the target boxes are regressed in the feature map of frame t. After regression, the classification score Skt is compared with the set threshold to determine whether to keep the target boxes. The second part is trajectory

initialization, as indicated by the red arrow in Fig. 1. For newly appeared trajectories, based on the detection results of the current frame provided by the target detector, the intersection over union (IoU) is calculated between the predicted bounding box set Bt and Dt. Bounding boxes with an IoU greater than the threshold new λ_{new} are filtered out, and the remaining bounding boxes are added as the first frame of the new trajectory to the active trajectories, forming the trajectory of the moving target.

In the above process, based on the bounding box frame of the video target, direct regression of the Tracktor algorithm should guarantee two assumptions: the camera does not have severe movement (or the construction personnel do not have severe movement); the video frame rate cannot be too low. These two points may not necessarily be satisfied in practice. Bergmann et al. [9] added two additional contents to form the Tracktor++ algorithm for this situation. First, a camera motion estimation compensation model based on Enhanced Correlation Coefficient (ECC) is used for camera movement compensation. In addition, a constant velocity model was adopted to calculate the position of time t based on the central coordinates of time t-2 and t-1, and then the Tracktor algorithm was used for regression. To reduce identity switches, the algorithm integrates a parallel re-identification network implemented by a Siamese CNN [15]. This component generates discriminative feature embeddings for personnel images and performs similarity measurement in the learned feature space to ensure target identity consistency throughout the tracking process.

2.2 Algorithm Feature Enhancement Tracktor++ Algorithm

Although the Tracktor++ algorithm has optimal design concepts, it cannot be directly applied to the electric power construction with changing environments and intense personnel movement. To enhance its performance in electric power construction, target improvements have been made to its target recognition network, motion compensation part, and re-identification network.

2.2.1 Attention Enhanced Object Detection Network

Regarding the specific task of locating and tracking personnel at power construction sites, the environmental background of personnel may vary significantly with seasons. Key factors such as lighting conditions, target sizes, and significant changes in personnel position and orientation angles all impose high demands on positioning and detection capabilities of the model. Tracktor++, which adopts the Faster-RCNN as the target detection part of the network structure, is not suitable for complex environments.

Without increasing the complexity of the model, the network should adaptively pay more attention to construction workers and improve the ability to locate targets. The main network in Faster-RCNN is first replaced with a more efficient Resnet-50 network, and then the feature pyramid structure (Feature

Pyramid Network, FPN) is introduced. In addition, a SimAM attention module is introduced at four output positions of the main network to optimize features extraction. The attention mechanism-enhanced object detection network after improvement is shown in Fig. 2. ***C*** represents the convolutional layer; ***M*** represents the merge layer, and ***P*** represents the predicted feature map.

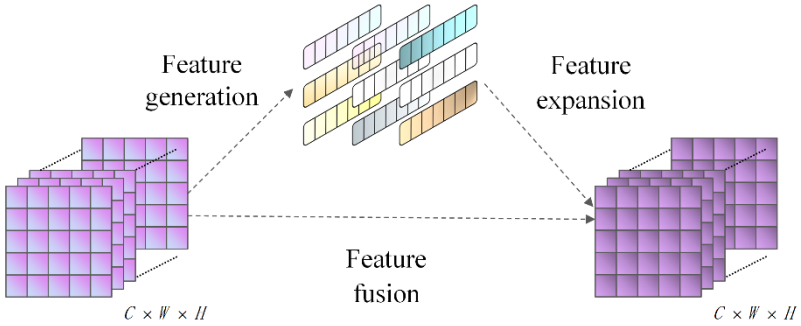

Fig. 2 The structure of the improved detection network with SimAM and FPN

The SimAM self-attention mechanism is used to enhances feature representation of small personnel targets and mitigates background interference, thereby improving personnel detection accuracy while preserving a certain level of detection efficiency. The SimAM structure is shown in Fig. 3.

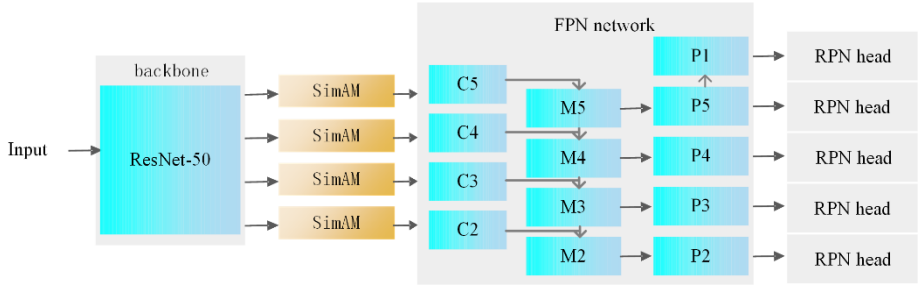

Fig. 3 The structure of SimAM

SimAM is established based on the visual neuroscience theory: neurons exhibiting spatial suppression effects are more critical than their neighboring neurons [16]. In personnel positioning tasks on construction sites, these neurons often extract key personnel features and require higher weights. The SimAM module without parameters is as follows:

$$Y = Sigmoid(\frac{1}{E(x)}) \square \ X \tag{1}$$

where $E(x)$ is the energy function of all channels. The design of the energy function inspired by neuroscience theory aims to quantify the linear separation between neurons, enabling the identification of critical neurons. For each input neuron, the

energy function e_t is defined as shown in Eq. 2. The smaller e_t, the higher the ability to distinguish of the neuron to its neighboring neurons.

$$\begin{aligned} e_t(x_t) &= \frac{4(\sigma^2+\lambda)}{(x_t-\mu)^2+2\sigma^2+2\lambda} \\ \mu &= \frac{1}{W\times H}\sum_{i=1}^{W\times H} x_i \\ \sigma^2 &= \frac{1}{W\times H}\sum_{i=1}^{W\times H}(x_i-\mu)^2 \end{aligned} \tag{2}$$

where x_t represents the target neuron of the input feature in the current channel. The average and variance of all neurons in this channel are set as μ and σ^2. The weight constant λ is set to 0.0001 [11].

The SimAM module is integrated at the connection between the backbone network and the FPN structure. Unlike embedding the SimAM module with the backbone network, placing it at this junction allows the network to fully leverage the pre-trained weights from large-scale datasets (such as ImageNet-1K [17] and MS COCO [18]), eliminating the need for extensive on-site construction data to learn human target features and significantly reducing training costs.

2.2.2 Kalman filter motion compensation

Tracktor++ employs the camera motion compensation (CMC) of the Enhanced Correlation Coefficient (ECC) method. However, commonly used safety monitoring cameras on construction sites are usually fixed on tripods, indicating camera movement is rare in such scenarios. On the other hand, construction workers have far more intense movement than ordinary individuals, resulting in the maximum frame rate of the camera being lower than the number of frames required for individuals.

Therefore, it is necessary to design a new motion estimation algorithm that takes into account the minimal camera displacement and rotation of significant and unpredictable personnel movement on construction sites. An eight-dimensional space is employed to describe the motion state of an object, derived from a positioning algorithm x= (u, v, γ, h, u', v', γ', h'). (u, v) is the center coordinates of the target bounding box. γ denotes the aspect ratio of various image objects, and h represents the target height. The remaining terms correspond to the time derivatives of the above quantities. A Kalman filter is then used to predict and update the object trajectory x' in frame t+1 based on Eq. 3.

$$\begin{aligned} x' &= \boldsymbol{F}x \\ \boldsymbol{F} &= \begin{bmatrix} \boldsymbol{E} & diag(\frac{d}{dt}) \\ 0 & \boldsymbol{E} \end{bmatrix} \end{aligned} \tag{1}$$

where x represents the state at time t, and $\boldsymbol{F}$ is the state transition matrix. We perform regression on the personnel positions in the inference-derived trajectory x' for frame t+1.

2.2.3 Re-identification network with the ResNeXt and batch normalization neck

Re-identification (Re-ID) networks, which enable differentiation between distinct individuals, are pivotal to tracking the movement trajectories of construction personnel. Their recognition performance directly determines the reliability of such tracking [19]. To enhance tracking effectiveness, it is essential to optimize Re-ID networks by leveraging the significant height variations of personnel in construction site videos, thereby strengthening their discriminative capability for individual differentiation.

To re-identify construction workers with different heights, algorithms should possess robust multi-layer feature fusion capabilities to capture multi-scale features of workers. The Tracktor++ algorithm uses the ResNet [20] network as the re-identification backbone, but its feature extraction capability is somewhat limited. To address this, we adopt ResNeXt [21] as the backbone for re-identification.

Compared with ResNet, ResNeXt has a deeper and wider architecture, and more effective target feature extraction without increasing hyper parameters. The basic building blocks of ResNeXt and ResNet are compared as shown in Fig. 4.

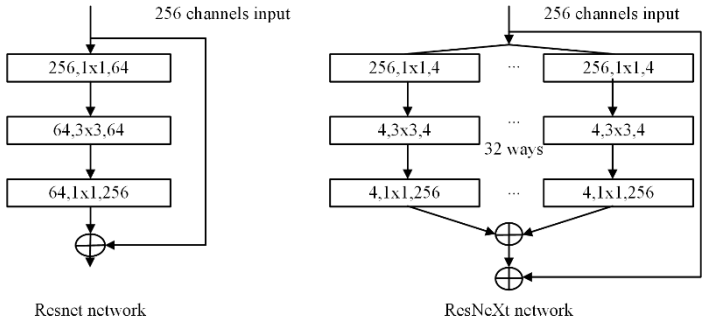

Fig. 4 Residual blocks of Resnet and ResNeXt

As shown in Fig. 4, ResNeXt first reduces the channel dimension of the input feature matrix from 256 to 128 via a 1×1 convolution. Then the feature map is split into 32 groups, each processed by a 3×3 convolution. The outputs of these groups are then concatenated to aggregate the extracted features. Finally, the concatenated features are up sampled back to the original dimension using a 1 × 1 convolution, and a residual connection is applied before the final output is activated by a ReLU function. The aggregation transformation of the ResNeXt network is as follows:

$$\boldsymbol{Y} = \boldsymbol{X} + \sum_{i=1}^{C} T_i(\boldsymbol{X}) \tag{2}$$

where, $\boldsymbol{Y}$ denotes the output; $\boldsymbol{X}$ denotes the input; C denotes the size of the aggregation transformation set, and $T_i(\boldsymbol{X})$ represents the convolution on $\boldsymbol{X}$. The layer structure of ResNeXt-50 is shown in Tab. 1, comprising a total of 5 convolutional layers. First, a 7×7 convolution is used for down-sampling in convolution layer 1, and then a 3×3 convolution kernel is used for down-sampling in convolution layer 2. These stacked blocks are then repeated 3, 4, 6, and 3 times respectively.

In the Tracktor++ algorithm, the re-identification network employs a linear recognition head (Liner Head) as the recognition head, as shown in Fig. 5(a). This head only consists of one or multiple fully connected layersaiming to transform feature vectors output by the backbone network into one-hot encoding. These encodings are then used to construct identity recognition loss values (such as cross-entropy loss). To optimize the model, identity loss and triplet loss are jointly employed.

Table 1

Structure of ResNeXt-50

Layers	Output size
Convolution layer 1	112×112
Max pooling layer	112×112
Convolution layer 2	56×56
Convolution layer 3	28×28
Convolution layer 4	14×14
Convolution layer 5	7×7
Mean pooling layer	1×1

SphereFace algorithm [22] and SphereReID [23] algorithm have shown that networks using linear detection heads may suffer from asynchronous convergence between identity recognition loss and triplet loss. Specifically, when one loss value continues to converge and decrease and the other first increases and then decreases at a certain stage, the gradient directions between the two tasks become inconsistent during updating. To address the above issues, the Batch Normalize Neck (BNNeck) [24] is used to constrain the feature map of the triplet loss in a free Euclidean space, and the identity recognition loss can be constrained to a hypersphere. The structure of BNNeck is shown in Fig. 5(b). The workflow of BNNeck is as follows: the network first calculates the triplet loss directly on the feature maps extracted from the backbone; then batch normalization is applied to make these features distribute approximately near the hypersphere; finally, the normalized feature vectors are input into the fully connected layer to compute the identity loss.

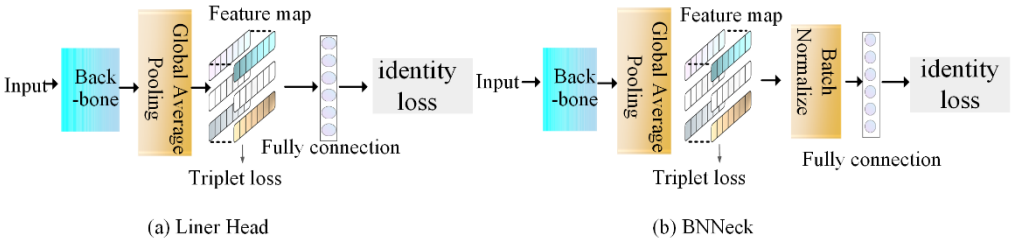

Fig. 5 The structure of Liner head versus the BNNeck head

The structure of the re-identification network after introducing ResNeXt and BNNeck is shown in Fig. 6. The newly detected and deactivated targets are respectively input into two ResNeXt-50 networks with shared weights. The output feature maps of these two networks are input into a BNNeck network to connect to a fully connected layer after batch normalization. This process generates one-hot encodings for the newly detected and deactivated targets. Cross-entropy loss is calculated based on these encodings to quantify the similarity score between the newly detected and deactivated targets. If the score is higher than the threshold, the new object is assigned the same identity number as the inactive object; otherwise, it uses a new number.

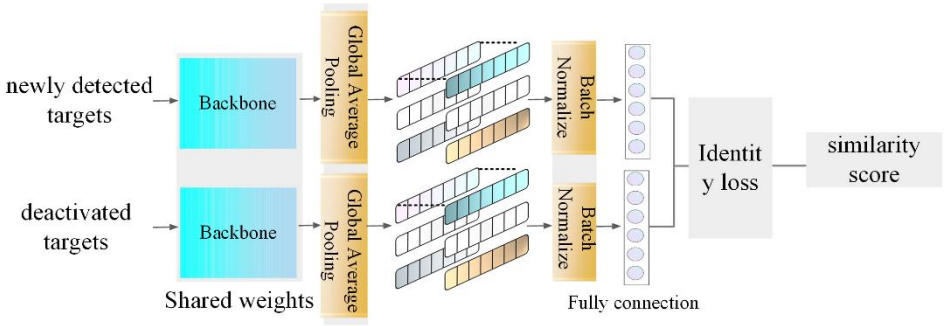

Fig. 6 The improved ReID network using ResNeXt backbone and BNNeck head

3. Results and analysis

3.1 Dataset

To validate the effectiveness of the F-Tracktor++ algorithm, a dataset for object recognition and tracking in power construction sites was collected using cameras deployed at the power construction site. The video frame rate is 25 frames per second (FPS), and the resolution is 1920x1080. Capturing scenarios from various construction sites, locations were carefully selected to align with typical camera positions and pitch angles in real-world construction. In addition, the videos were filmed under different weather conditions to obtain a wide range of

samples. The dataset is publicly available in https://github.com/zhanghay/POWER-CONSTRUCTION-SITE-.

The annotation software is used to annotate the positions and identity the information of construction personnel frame by frame. Finally, according to the file organization style of the MOT-17 [25] dataset, the construction site personnel tracking dataset is formed, as shown in Fig. 7.

3.2 Training Settings

The experiment runs on the Ubuntu 22.04 LTS operating system by using two NVIDIA RTX 1080Ti GPUs. The pytorch 1.10.1 and CUDA 11.3 are used as the deep learning framework.

Fig. 7 Four example screenshots of the on-site dataset for power construction

The weights of the target recognition network are first trained. The backbone network adopts the parameters of the pre-trained ResNet-50 network in the MS COCO dataset and fine-tuned on the construction site dataset. For fine-tuning, the first residual layer is frozen, and the mean and variance of the normalization layer are fixed. The classification loss in the RPN detection head uses cross-entropy loss, and the bounding box loss uses smooth L1 loss. Both loss values have a weight of 1.0. The fully connected layer of the ROI detection head has 1024 nodes and outputs 1 classification. The classification loss uses cross-entropy loss, and the target bounding box loss uses smooth L1 loss. The weights of the two losses are both 1.0. Then, the weight of the re-identification network is trained. ResNeXt-50 inherits from the ImageNet-1K pre-trained model, which contains 4 residual modules and selects the last layer as the output. The fully connected layer of BNNeck has 1024 nodes, and the output channel is 128. The loss function uses cross-entropy loss, and the activation function uses ReLU activation function.

Fig. 8(a) shows the training loss curve of the target detection network, and Fig. 8(b) shows the curve of the loss function of the re-identification network. As shown in Fig. 8, both networks have converged to a stable state after training.

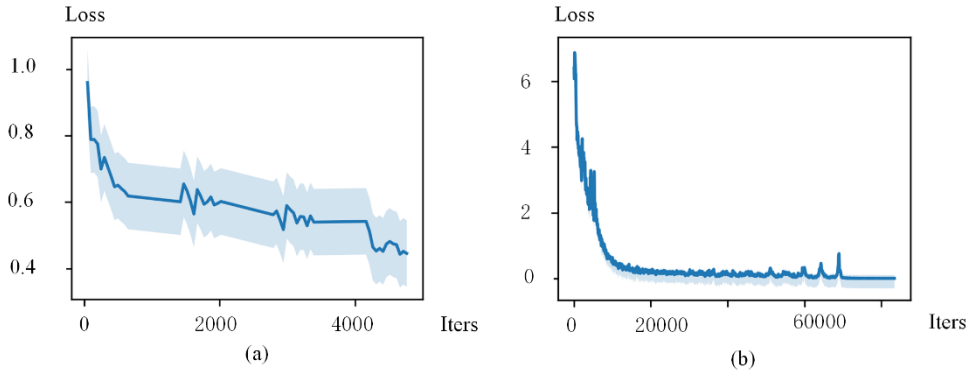

Fig. 8 Training loss curves

3.3 Experimental Indicators

To track multiple construction personnel, it is necessary to use multi-target tracking evaluation indicators to measure the tracking performance of multi-target tracking algorithms. These metrics include: False Positive (FP), which means the target is not in the ground truth but in the trajectory generated by the tracking algorithm; False Negative (FN), the target is in the ground truth but there is no matching target in the results of the tracking algorithm; total number of identity switches (IDs): the number of times that the same target is identified with different IDs; Mostly Tracked (MT): the number of trajectories where 80% of the targets are successfully tracked; Mostly Lost (ML): the number of trajectories where at least 80% of the targets are lost. Identification Precision (IDP) refers to the accuracy of identifying the identity in the target box, which is the proportion of samples identified as positive that are actually positive.

3.4 Experimental Results and Discussion

To verify the effect of the improved strategy on the Tracktor++ algorithm, the following ablation experiments are designed after applying different improvement measures. The results of the ablation experiments are shown in Tab. 2.

Table 2

Results of ablation experiment

Methods	FP	FN	IDs	MT	ML	IDP
Tracktor++	3346	3954	664	232	28	0.57
Use feature enhanced detection network	3240	3899	664	234	26	0.57
Use motion estimation with Kalman filte	3239	3895	664	234	26	0.57
F-Tracktor++	3239	3895	582	234	26	0.65

As shown in Table 2, adding the attention module to the network can reduce the false positive and negative values, indicating more accurate target recognition. This can reduce the probability of missing and false alarms, as well as slightly increasing the success rate of target tracking. On this basis, introducing Kalman filter as motion estimation can slightly improves false positives and negatives. The improvement in re-identification can reduce the total number of identity exchanges and increase the accuracy of identity recognition. This indicates that the improved Tracktor++ module demonstrates enhanced precision in personnel identity discrimination. Compared with the Tracktor++ algorithm, the comprehensive optimization strategy achieves a 3.2% reduction in the overall false positive rate, a 12.3% decrease in the total identity switches, and a 90% success tracking rate. Based on the improved accuracy of personnel identification and recognition, the accuracy of tracking trajectory is ensured, which helps to improve the accuracy of personnel positioning and tracking in construction sites.

The F-Tracktor++ algorithm is compared with the UWA [26] algorithm and the DeepSORT [13] algorithm, as shown in Tab. 3.

Table 3

Result of comparative experiments

Methods	FP	FN	IDs	MT	ML	IDP
Tracktor++	3346	3954	664	232	28	0.57
UMA	3540	4299	973	134	126	0.34
DeepSORT	3831	4271	831	221	39	0.42
F-Tracktor++	3239	3895	582	234	26	0.65

Fig. 9 shows the localization and tracking results in the test set, with construction workers marked by target bounding boxes. The numbers in the top left corner of the bounding boxes indicate personnel identity and confidence. The point-like marks represent the movement trajectories of each construction worker. The dataset includes instances of partial occlusion and person crossing, where the occlusion area is relatively small (occlusion area < 20%), as shown in Fig. 9, and the occlusion duration is brief (occlusion duration < 3 s). Under these conditions, the proposed algorithm is robust and can achieve accurate target identification and tracking.

Fig. 9 Some tracking results in the test dataset

Since the original dataset lacked variations in illumination, a low-brightness subset was synthesized using image processing techniques. Brightness and contrast were artificially reduced in the test set to generate a low-light subset (representative frames shown in Fig. 10). When evaluated on this subset, the F-Tracktor++ algorithm achieved the following metrics: FP = 3265, FN = 3926, IDs = 593, MT = 233, ML = 27, and IDP = 0.62. The results indicate that compared with normal lighting conditions, the recognition accuracy decreased. The number of identity switches (IDs) significantly increased by 9, indicating that low-light conditions will significantly affect the reliability of target identification.

To evaluate the industrial applicability of the proposed algorithm, especially its real-time diagnostic capability, the algorithm model is converted to a TensorRT format and deployed on the experimental setup, as shown in Fig. 11. It employs an NVIDIA Jetson NX as the computing device for inference. The experimental results show that the inference speed of the device is 16 Frames Per Second (FPS), meeting the basic requirements of real-time processing.

Fig. 10 Example of low brightness dataset images

Fig. 11 Equipment used for testing

4. Conclusion

A construction site video dataset is established. Based on the original Tracktor++ framework, the SimAM attention module is introduced to further improve the feature extraction and fusion capabilities of target recognition. The Kalman filter is introduced in the motion estimation to improve the tracking ability of moving personnel. To address the requirements of personnel identity feature extraction, ResNeXt is adopted as the backbone network, and BNNeck serves as the detection head for the re-identification (Re-ID) module. Subsequently, transfer learning and optimized training strategies are employed to accelerate model convergence, ensuring efficient training without compromising performance. In testing, the improved algorithm outperforms the original Tracktor++ by a 3.2% reduction in the overall false positive rate, a 12.3% decrease in total identity switches, and a 90% successful tracking rate. Based on the improvement in the accuracy of personnel identification and recognition, the accuracy of tracking trajectories is guaranteed. Furthermore, experimental validation demonstrates that the proposed algorithm achieves a 16 FPS frame rate on edge computing devices, meeting the requirements of real-time industrial diagnostics.

The comprehensive experimental results show that the improved Tracktor++ algorithms perform well in personnel positioning and tracking on the electric power construction site. There is still room for improvement in identifying smaller targets and tracking obscured individuals. This is may be limited by the structure of the Faster-RCNN target recognition network, and further research would focus on addressing this problem.

Acknowledgement

This study is funded by science and technology project of Hebei Electric Power Company of State Grid Corporation of China (kj2022-009).

REFERENCES

[1] *Yang, Y., Bo, B., Zheng, Y., Lu, Y., Zhou, B., Ji, H., & Shen, D.,* Research on power engineering solutions based on NB-IoT technology. 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA), Shenyang, China, IEEE, pp. 224-228, 2022. Doi: 10.1109/ICPECA53709.2022.9718830.

[2] *Peng, G., Lei, Y., Li, H., Wu, D., Wang, J., & Liu, F.,* CORY-Net: Contrastive Res-YOLOv5 Network for Intelligent Safety Monitoring on Power Grid Construction Sites. IEEE Access, vol. 9, pp. 160461-160470, 2021. Doi: 10.1109/ACCESS.2021.3132301

[3] *Bartenev, D. A., Telyatnikova, N. A., Prokofieva, V. N., & Mahendra, A. J.,* Development of Solutions to Increase the Power Supply Efficiency in the Electrical Management of Construction Site. 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus), Petersburg, Moscow, Russia, IEEE, pp. 1874-1878, 2021. Doi: 10.1109/ElConRus51938.2021.9396490

[4] *Madavarapu, J. B., Nachiyappan, S., Rajarajeswari, S., Anusha, N., Venkatachalam, N., Madavarapu, R. C. B., & Ahilan, A.,* HOT Watch: IoT-Based Wearable Health Monitoring System. IEEE Sensors Journal, vol. 24, no. 20, pp. 33252-33259, 2024. Doi: 10.1109/JSEN.2024.3424348

[5] *Sarmento, J., Dos Santos, F. N., Aguiar, A. S., Sobreira, H., Regueiro, C. V., & Valente, A.,* FollowMe - A Pedestrian Following Algorithm for Agricultural Logistic Robots. 2022 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC), Maria da Feira, Portugal, IEEE, pp. 179-185, 2022. Doi: 10.1109/ICARSC55462.2022.9784791

[6] *Yang, J., Zhang, S., Wang, M., & Liu, W.,* Human motion trajectory generation based on AWR1642. 2023 International Conference on Computers, Information Processing and Advanced Education (CIPAE), Ottawa, ON, Canada, IEEE, pp. 178-181, 2023. Doi: 10.1109/CIPAE60493.2023.00040

[7] *Yang, B., Li, J., Shao, Z., & Zhang, H.,* Self-Supervised Deep Location and Ranging Error Correction for UWB Localization. IEEE Sensors Journal, vol. 23, no. 9, pp. 9549-9559, 2023. Doi: 10.1109/JSEN.2023.3258432

[8] *Miao, J., Chunyang, J., Yanli, Z., Zhenyu, J., & Wei, G.,* Safety Detection Method for Work at Height Based on Improved YOLOv8. 2024 10th International Symposium on System Security, Safety, and Reliability (ISSSR), Xiamen, China, IEEE, pp. 206-210, 2024. Doi: 10.1109/ISSSR61934.2024.00031

[9] *Bergmann, P., Meinhardt, T., & Leal-Taixe, L.,* Tracking without bells and whistles. 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), IEEE, pp. 941-951, 2019. Doi: 10.1109/iccv.2019.00103

[10] *Ren, S., He, K., Girshick, R., & Sun, J.,* Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp.1137-1149, 2016. Doi: 10.1109/tpami.2016.2577031.

[11] *Yang, L., Zhang, R. Y., Li, L., & Xie, X.,* Simam: A simple, parameter-free attention module for convolutional neural networks. International conference on machine learning, pp. 11863-11874, 2021. Doi: 10.36227/techrxiv.22662601.v1

[12] *Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B.,* Simple online and realtime tracking. 2016 IEEE international conference on image processing (ICIP), Phoenix, AZ, USA, IEEE, pp. 3464-3468, 2016. Doi: 10.1109/ICIP.2016.7533003

[13] *Pujara, A., & Bhamare, M.,* Deepsort: real time & multi-object detection and tracking with YOLO and TensorFlow. 2022 International Conference on Augmented Intelligence and Sustainable Systems (ICAISS), Trichy, India, IEEE, pp. 456-460, 2022. Doi: 10.1109/icaiss55157.2022.10011018

[14] *Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., & Meng, H.,* Strongsort: Make deepsort great again. IEEE Transactions on Multimedia, vol. 25, pp.8725-8737, 2023. Doi: 10.1109/tmm.2023.3240881

[15] *Koch, G., Zemel, R., & Salakhutdinov, R.,* Siamese neural networks for one-shot image recognition. ICML deep learning workshop, vol. 2, no. 1, pp. 1-30, 2015. Doi: 10.1109/icacite57410.2023.10182466

[16] *Li, Z.M., Li, J.H., Yin, F., Tie, J., Wu, Q.B.,* Dense citrus detection algorithm based on deformable convolution and SimAM attention. Journal of Chinese Agricultural Mechanization, vol. 44, no. 2, pp.156-162, 2023. Doi: 10.13733/j.jcam.issn.2095-5553.2023.02.022

[17] *Deng, J., Dong, W., Socher, R., Li, L.J., Li, K. and Li, F.F.,* Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition, Miami, FL,

USA, IEEE, pp. 248-255, 2009. Doi: 10.1109/cvprw.2009.5206848

[18] *Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P. and Zitnick, C.L.,* Microsoft coco: Common objects in context. Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014. Proceedings, Part, vol. 13, pp. 740-755, 2014. Doi: 10.1007/978-3-319-10602-1_48

[19] *Jung, H., Kang, S., Kim, T. and Kim, H.,* ConfTrack: Kalman filter-based multi-person tracking by utilizing confidence score of detection box. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 6583-6592, 2024. Doi: 10.1109/wacv57701.2024.00645

[20] *He, K., Zhang, X., Ren, S. and Sun, J.,* Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, IEEE, pp. 770-778, 2016. Doi: 10.1109/cvpr.2016.90

[21] *Xie, S., Girshick, R., Dollár, P., Tu, Z., & He, K.,* Aggregated residual transformations for deep neural networks. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, IEEE, pp. 1492-1500, 2017. Doi: 10.1109/cvpr.2017.634

[22] *Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L.,* Sphereface: Deep hypersphere embedding for face recognition. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, IEEE, pp. 212-220, 2017. Doi: 10.1109/cvpr.2017.713

[23] *Fan, X., Jiang, W., Luo, H., & Fei, M.,* Spherereid: Deep hypersphere manifold embedding for person re-identification. Journal of Visual Communication and Image Representation, vol. 60, pp.51-58, 2019. Doi: 10.1016/j.jvcir.2019.01.010

[24] *Luo, H., Jiang, W., Gu, Y., Liu, F., Liao, X., Lai, S., & Gu, J.,* A strong baseline and batch normalization neck for deep person re-identification. IEEE Transactions on Multimedia, vol. 22, no. 10, pp.2597-2609, 2019.Doi: 10.1109/tmm.2019.2958756

[25] *Dendorfer, P., Osep, A., Milan, A., Schindler, K., Cremers, D., Reid, I., Roth, S. and Leal-Taixé, L.,* Motchallenge: A benchmark for single-camera multiple target tracking. International Journal of Computer Vision, vol. 129, pp.845-881, 2021. Doi: 10.1007/s11263-020-01393-0

[26] *Yin, J., Wang, W., Meng, Q., Yang, R., & Shen, J.,* A Unified Object Motion and Affinity Model for Online Multi-object Tracking. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, IEEE, pp. 6768-6777, 2020. Doi:10.1109/CVPR42600.2020.00680