Heliyon 10 (2024) e26184

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

An effective electricity worker identification approach based on Yolov3-Arcface

Qinming Liu, Fangzhou Hao, Qilin Zhou, Xiaofeng Dai, Zetao Chen, Zengyu Wang *

Tianhe Power Supply Bureau of Guangzhou Power Supply Bureau, Guangdong Power Co. Ltd., Guangzhou, 510000, China

ARTICLE INFO

Keywords:
Power distribution room
Face recognition
YOLOv3
ArcFace
Detection performance

ABSTRACT

To address the issues of low efficiency and high complexity of detection models for electric power workers in distribution rooms, the electric power worker identification approach is proposed. The ArcFace loss function is used as the coordinate regression loss of the target box. According to the score, the template box with the highest score is selected for prediction, which speeds up the rate of convergence. Dimensional clustering is used to set template boxes for bounding box prediction. The experimental results show that the improved YOLOv3 is a high-performance and lightweight model. The electric power worker identification approach proposed in this paper has a high-speed recognition process, accurate recognition results. The effectiveness of the approach is verified with better detection performance and robustness.

1. Introduction

The production site of electric power workers has very high requirements for personnel safety management [1–3]. Although some units have implemented door security, it is generally applied in a small range, most of which are only for office areas [4–8]. At the same time, there is a common situation of mixed traffic between people and vehicles, and related statistics rely on manual labor. There are many human factors, and statistical deviations are large, posing safety hazards [9,10]. Due to the complex environment of the power distribution room and the significant changes in facial angles of power workers during rapid movement, resulting in blurred movement, if the identity of power workers is not accurately identified, it will affect the safety of the power distribution room [11,12]. Therefore, to improve the security management of the distribution room, it is necessary to identify and alert unrelated personnel entering the distribution room.

Deep learning technology in artificial intelligence fields such as computer vision, face recognition, as an important component of computer vision technology, has also made significant progress [13–15]. It has stronger robustness and generalization ability compared to traditional detection algorithms [16,17]. Deep learning based face recognition methods extract generalized facial feature representations by training on massive face data, highlighting the distinguishability between features, which greatly improves detection accuracy [18]. Reference [19] proposes a DeepFace based facial recognition model that uses a single unmodified 3D surface face as an approximation of the input facial image. Reference [20] applied triple loss to the VGG model, and experimental results showed that using triple loss instead of Softmax loss can further improve recognition accuracy. Facial features learned by CNN (Convolutional Neural Network) networks, reference [21] proposed a face recognition model based on Center loss. Center loss learns the center point features of each category, and sets a penalty function to improve the recognition ability of the learned features.

* Corresponding author.
E-mail address: wangzengyugzps@gmail.com (Z. Wang).

https://doi.org/10.1016/j.heliyon.2024.e26184
Received 12 May 2023; Received in revised form 30 January 2024; Accepted 8 February 2024
Available online 14 February 2024

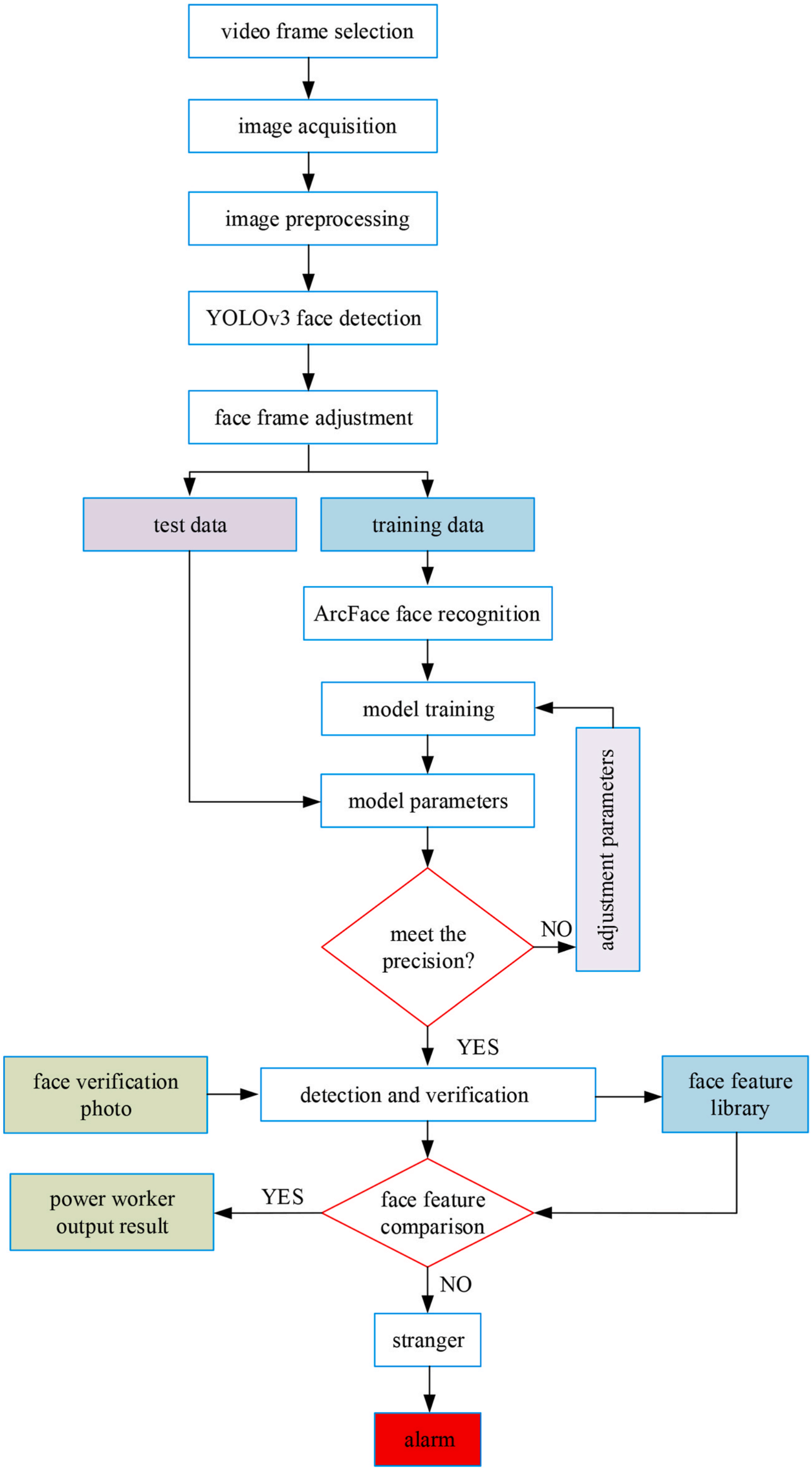

Fig. 1. The flowchart of the identification system for electric power workers.

Compared with AM softmax loss function, which maximizes the classification boundary in cosine space, ArcFace Loss function proposed in Ref. [22] maximizes the classification boundary in angle space. Compared with AM softmax and other loss function, ArcFace has stronger geometric interpretation and stricter requirements for classification. Literature [23] combines classification and

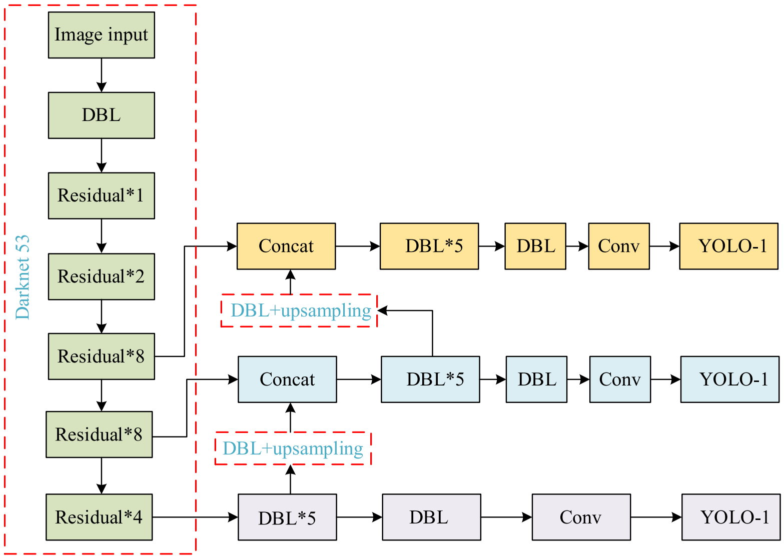

Fig. 2. YOLOv3 network structure.

localization tasks to obtain the position and category of targets after a single feature extraction. However, it is not sensitive enough and has a low accuracy when detecting small-sized targets.

In response to the multi-scale problem, the YOLOv3 at the corresponding position of the target to perform regression processing on its bounding box, reducing computational complexity and making the design relatively simple. Reference [24] applied the original version of YOLOv3 algorithm to the field of real-time gesture recognition in experiments, and although it achieved good results with high accuracy, the training time is too long [25]. The network structure based on YOLOv3 algorithm in Ref. [26] uses ResNet101 instead of DarkNet-53 feature extraction network to improve detection performance. Reference [27] used the Jaccard index as the evaluation and detection indicator, used the Soft IoU layer for candidate box prediction, and proposed the EM Merge unit to solve the problem of candidate box overlap. Reference [28] proposes a single stage detection algorithm for dense targets.

An electric power worker identity recognition method based on YOLOv3 is proposed. Based on the YOLOv3 algorithm, facial detection is performed on electric power workers, and the detected facial frames are then fed into the ArcFace loss architecture. After the electric power worker applies for two ticket permissions, they need to enter the power distribution room for work operations, and the facial recognition model identifies the operator's identity. Implement identity recognition and alarm functions through facial recognition technology to prevent unnecessary risks from other unrelated personnel entering the electrical room.

2. Identification framework for electric power workers

Power worker identification mainly includes two parts: face frame detection and face recognition [29]. Face frame detection is mainly based on the image target detection model of deep learning; face recognition is the process of matching the feature vector value of the face in the face frame detected by the target and the full feature vector stored in the database. First, perform face detection on the input. If a face is detected, output the coordinates of the face frame. Then perform face alignment, and combine the face frame output by face detection to detect the positions of the eye, nose, and mouth feature points, and the model outputs the coordinates [30]. Then, input the ArcFace model for feature extraction, fine-tune its model parameters. By comparing the distance and similarity of the two vectors, the analysis results are retrieved and matched in the face information management system, and the retrieved personnel information is returned to the front-end page to help users quickly identify personnel [31–33]. YOLO adopts the idea of regression, using the entire image as input to CNN, and then directly regressing the object's bounding box and category labels in the output layer. The framework is shown in Fig. 1.

3. Electricity worker identification approach based on YOLOv3-ArcFace

3.1. Data preprocessing

Firstly, the processed strings are used as the input of the face recognition service. Then face detection is performed on the input, and the identity of the worker is extracted through Haar-like features, and the 105 key points of face recognition are marked (24 points for eyebrows, 32 points for eyes, 6 points for nose, 34 points for mouth, 9 points for outer contour).

It is necessary to perform face alignment and normalization on the face image. By extracting feature points of the orbit of the face, the center position of eyes can be obtained. According to the center position of the two eyes, the deflection angle can be determined as [7]:

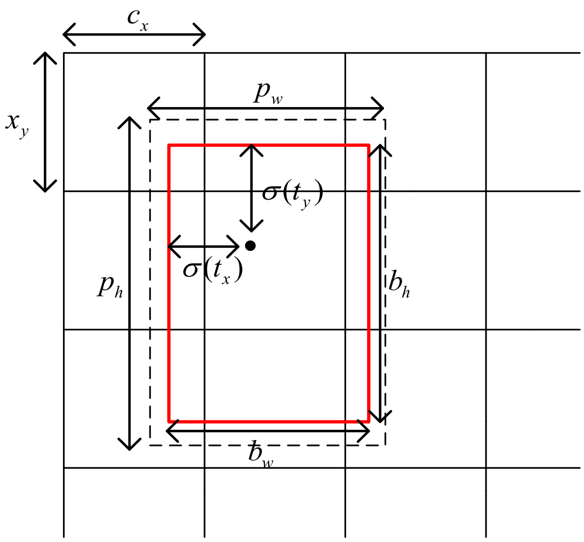

Fig. 3. YOLOv3 bounding box prediction.

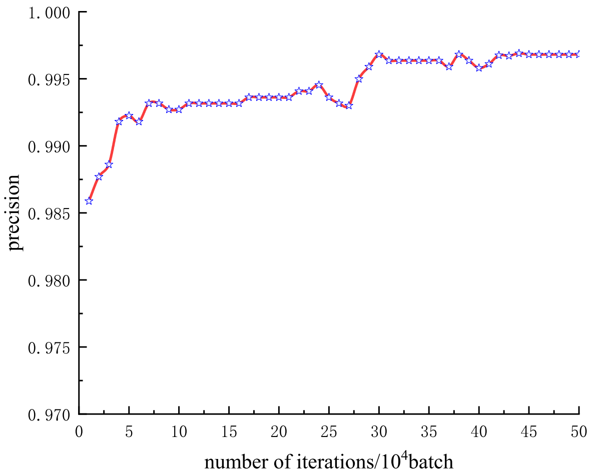

Fig. 4. Training curve using ArcFace loss.

$$\alpha = \arctan\left(\frac{|y_l - y_r|}{|x_l - x_r|}\right) \tag{1}$$

Where, y_l and y_r are vertical coordinate; x_l and x_r are horizontal coordinate.

3.2. Yolov3 face detection

YOLO algorithm is suitable for practical engineering applications. The DarkNet-53 network structure further improves the recognition accuracy of YOLOv3 by deepening the network structure while ensuring the real-time detection of YOLO. This paper uses YOLOv3 as the basic framework of the face detection module. The loss function can be described as [15]:

$$\text{loss} = -\frac{1}{n}\sum_x [y\ln(a) + (1-y)\ln(1-a)] \tag{2}$$

Where, n is sample size; x is the sample; a is the probability of information entropy; y is the probability of relative entropy.

The proposed structure is shown in Fig. 2. The feature maps are up-sampling and tensor splicing twice respectively, and then outputs at three different scales are obtained. Using multi-scale to perform face detection on targets of different sizes, even if the target face is small, it can be successfully detected, which improves the prediction accuracy.

Table 1
Darknet53 network.

	type	convolution information	feature map size
Res1	convolutional	32 3*3	416*416
		64 3*3/2	208*208
		32 1*1	–
		64 3*3	–
	residual	–	208*208
Res2	convolutional	128 3*3/2	104*104
		64 1*1	–
		128 3*3	–
	residual	–	104*104
Res8	convolutional	256 3*3/2	52*52
		128 1*1	–
		256 3*3	–
	residual	–	52*52
Res8	convolutional	512 3*3/2	26*26
		256 1*1	–
		512 3*3	–
	residual	–	26*26
Res4	convolutional	1024 3*3/2	13*13
		512 1*1	–
		1024 3*3	–
	residual	–	13*13

Dimensional clustering is used to set template boxes for bounding box prediction. Through the relevant variable, the coordinates of the center point can be shown in Fig. 3. Table 1 presents the Darknet53 network.

3.3. Face recognition based on ArcFace

The calculation formula of the ArcFace loss function is expressed as [21]:

$$\begin{cases} L_{\text{ArcFace}} = -\frac{1}{m}\sum_{i=1}^{m}\log\frac{e^{s\cos\left(\theta_{y_i}+m\right)}}{e^{s\cos\left(\theta_{y_i}+m\right)}+\sum_{j\neq y_i}e^{s\cos\left(\theta_j\right)}} \\ s.t.\quad \cos\theta_j = W_j^T x_i, W_j = \frac{W_j}{\|W_j\|}, x_i = \frac{x_i}{\|x_i\|} \end{cases} \tag{3}$$

Where, m is the number of features of the loss function; θ_{yi} is the predict values; W_j is the value of the model's true box; x_i is the category probability.

In (8), to introduce the angle loss function and eliminate the variation in the radial direction in the angle space, a normalization strategy needs to be adopted. Unlike the A-softmax loss function, which only normalizes the weight feature vector, the ArcFace loss function normalizes the weight feature vector.

Considering the case where the network initializes the ArcFace loss function at the beginning, normalize the features is a necessary step when the network is trained from scratch. The scaling parameter s also needs to be fixed instead of letting the network learn this parameter during the training phase. The value of the scaling parameter s is set to a larger integer, so that the network obtains a smaller training loss value and better performance. In the NSL loss function, the training loss value decreases as the value of s increases, but s is too small to make the network unable to converge. Therefore, in the ArcFace loss function, it is necessary to assign a large value to s to ensure that the network still has enough hypersphere space for feature learning when the margin value is large. The minimum boundary corresponding to the parameter s is [9]:

$$s \geq \frac{C-1}{C}\log\frac{(C-1)P_W}{1-P_W} \tag{4}$$

Based on the above boundaries, if the number of given categories is constant, the value of parameter s needs to be expanded to get the optimal P_W. If the number of categories increases when is P_W fixed, the value of parameter s should also increase accordingly. Because of the increase of categories, it increases the difficulty of multi-classification tasks in a relatively compact space.

4. Experimental verification and analysis

4.1. Environment configuration

The experimental parameters are shown in Table 2, and the training hyperparameter settings are shown in Table 3. At the beginning of training, adjust the rate to 0.01 after 10,000 iterations, 0.001 after 30,000 iterations, and 0.0001 after 40,000 iterations to

Table 2
Experimental environment.

Soft and hard items	Details
MEM	256 G
hard disk	>= 500G
operating system	Ubuntu18.04
Docker version	Docker version 19
GPU	NVIDIA TESLA P4
CUDA version	CUDA 10.1
Deep Learning Framework	Tensorflow2.x, Keras

Table 3
Training hyperparameter settings.

Hyperparameter	Value
Momentum	0.9
Learning rate	0.00125
decay	0.0005
Batchsize	32
Epoch	200~500

Table 4
Bounding box distribution.

Down sampling magnification	32	16	8
Feature map grid	13*13	26*26	52*52
visual perception	big	middle	small
a priori box	(116*90)(156*198)(373*326)	(30*61)(62*45)(59*119)	(10*13)(16*30)(33*23)

According to the prior frame settings of the three scale feature maps, there are a total of $13*13*3 + 26*26*3 + 52*52*3 = 10647$ a priori frames. The accuracy of validation set trained with ArcFace loss is shown in Fig. 4. When the network is trained to 160,000 steps, the network has basically converged, and the training is stopped. At this time, the accuracy is 99.67%.

Table 5
Performance comparison under different models.

Model	mAP	F1/%
Proposed algorithm	0.892	86.56
Adaboost	0.845	81.32
Faster-RCNN	0.881	79.58
PCA-SVM	0.834	84.36

Table 6
Accuracy comparison under different models.

Model	Wider-Face	Yale
Proposed algorithm	90.3%	96.7%
Adaboost	84.5%	87.3%
Faster-RCNN	87.8%	89.4
PCA-SVM	85.4%	87.1%

In Tables 5 and 6, in the Yale database, the recognition effect is the best, reaching 96.7%, which is higher than the traditional Adaboost method, Faster-RCNN and PCA-SVM methods are 9.6%, 7.5% and 9.9% higher, respectively, and the recognition effect is better. The F1 index in this paper is 86.56%, and its comprehensive performance surpasses other object detection algorithms.

further converge the loss function.

4.2. Experimental results

When the input is 416×416, the prior box distribution is shown in Table 4. A fully convolutional network with a stride of 2 is used to implement downsampling. In order to use deep-level features, the feature maps downsampled by 32 times and 16 times are upsampled. A new feature map with double the size is obtained, which becomes the dimension of 16 times and 8 times downsampling.

Conducted with the latest relevant literature methods (Adaboost [21], Faster RCNN [10], PCA-SVM [28]). The comparison of *mAP* and F1 indicators [34] under the same parameters is shown in Table 5. When using ORL face database and Yale face database [35,36], the comparison of recognition accuracy under different algorithms is shown in Table 6.

5. Conclusion

In order to prevent other unrelated personnel from entering the electric room generating unnecessary risks, an electric power worker identification method is proposed. Modification of loss function and regression function, the power worker identification model has been optimized. Compared with the Adaboost, Faster-RCNN and PCA-SVM methods, the accuracy of this paper under the Yale database is 9.6%, 7.5% and 9.9% higher, respectively, and the recognition effect is better. The improved algorithm meets the requirements of detection accuracy as well as real-time requirements, which is meaningful for better engineering applications.

In future work, we will construct specific facial datasets for training, while designing automatic filtering algorithms to automatically remove non-human facial targets, fuse multi-source information, and fuse information from multiple sources (such as action recognition and behavior prediction) to reduce error rate and enhance the superiority of the algorithm.

Data availability statement

The data that support the findings of this study are available at http://doi.org/10.1039/A708748I. and http://doi.org/WOS:000255193800030.

CRediT authorship contribution statement

Qinming Liu: Writing – review & editing, Writing – original draft, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Fangzhou Hao:** Writing – review & editing, Writing – original draft, Validation, Software, Methodology, Formal analysis, Data curation, Conceptualization. **Qilin Zhou:** Writing – review & editing, Writing – original draft, Methodology, Formal analysis, Data curation, Conceptualization. **Xiaofeng Dai:** Writing – review & editing, Writing – original draft, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Zetao Chen:** Writing – review & editing, Writing – original draft, Software, Formal analysis, Data curation, Conceptualization. **Zengyu Wang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Investigation, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is supported by Guangzhou Power Supply Company Project of Southern Power Grid, Project number (082400KK52180008).

References

[1] M. Sadeghi-Yarandi, S. Torabi-Gudarzi, N. Asadi, H. Golmohammadpour, V. Ahmadi-Moshiran, M. Taheri, B. Alimohammadi, Development of a novel Electrical Industry Safety Risk Index (EISRI) in the electricity power distribution industry based on fuzzy analytic hierarchy process (FAHP), Heliyon 9 (2) (2023) e13155, https://doi.org/10.1016/j.heliyon.2023.e13155.

[2] S.C. Nwanya, C.A. Mgbemene, C.C. Ezeoke, O.C. Iloeje, Total cost of risk for privatized electric power generation under pipeline vandalism, Heliyon 4 (7) (2018) e00702, https://doi.org/10.1016/j.heliyon.2018.e00702.

[3] J. Van der Laak, G. Litjens, F. Ciompi, Deep learning in histopathology: the path to the clinic, Nat. Med. 27 (5) (2021) 775–784, https://doi.org/10.1038/s41591-021-01343-4.

[4] K. Atz, F. Grisoni, G. Schneider, Geometric deep learning on molecular representations, Nat. Mach. Intell. 3 (12) (2021) 1023–1032, https://doi.org/10.1038/s42256-021-00418-8.

[5] H. Zhang, H. Xu, X. Tian, et al., Image fusion meets deep learning: a survey and perspective, Inf. Fusion 76 (2021) 323–336, https://doi.org/10.1016/j.inffus.2021.06.008.

[6] D. Brzezinska, P. Bryant, Performance-based analysis in evaluation of safety in car parks under electric vehicle fire conditions, Energies 15 (2) (2022) 649, https://doi.org/10.3390/en15020649.

[7] Y. Tian, G. Yang, Z. Wang, et al., Apple detection during different growth stages in orchards using the improved YOLO-V3 model, Comput. Electron. Agric. 157 (14) (2019) 417–426, https://doi.org/10.1016/j.compag.2019.01.012.

[8] S. Zhai, D. Shang, S. Wang, et al., DF-SSD: an improved SSD object detection algorithm based on DenseNet and feature fusion, IEEE Access 8 (4) (2020) 24344–24357, https://doi.org/10.1109/ACCESS.2020.2971026.

[9] T. Kong, F. Sun, H. Liu, et al., Foveabox: beyound anchor-based object detection, IEEE Trans. Image Process. 29 (14) (2020) 7389–7398, https://doi.org/10.1109/TIP.2020.3002345.

[10] S. Bian, C. Li, Y. Fu, et al., Machine learning-based real-time monitoring system for smart connected worker to improve energy efficiency, J. Manuf. Syst. 61 (2021) 66–76, https://doi.org/10.1016/j.jmsy.2021.08.009.

[11] E.A. Fadeyi, Emery, et al., Implementation of a new blood cooler insert and tracking technology with educational initiatives and its effect on reducing red blood cell wastage, Transfusion 57 (10) (2017) 2477–2482, https://doi.org/10.1111/trf.14234.

[12] X. Hua, Y. Ono, L. Peng, et al., Target detection within nonhomogeneous clutter via total bregman divergence-based matrix information geometry detectors, IEEE Trans. Signal Process. 69 (5) (2021) 4326–4340, https://doi.org/10.1109/TSP.2021.3095725.
[13] L. Tian, Y. Cao, B. He, et al., Image enhancement driven by object characteristics and dense feature reuse network for ship target detection in remote sensing imagery, Rem. Sens. 13 (7) (2021) 1327–1336, https://doi.org/10.3390/rs13071327.
[14] Z. Cao, X. Kong, Q. Zhu, et al., Infrared dim target detection via mode-k1k2 extension tensor tubal rank under complex ocean environment, ISPRS J. Photogrammetry Remote Sens. 181 (13) (2021) 167–190, https://doi.org/10.1016/j.isprsjprs.2021.09.007.
[15] X. Ren, C. Yue, T. Ma, et al., Adaptive parameters optimization model with 3D information extraction for infrared small target detection based on particle swarm optimization algorithm, Infrared Phys. Technol. 117 (3) (2021) 103838–103847, https://doi.org/10.1016/j.infrared.2021.103838.
[16] A. Safonova, E. Guirado, Y. Maglinets, et al., Olive tree biovolume from UAV multi-resolution image segmentation with mask R-CNN, Sensors 21 (5) (2021) 1617–1628, https://doi.org/10.3390/s21051617.
[17] F. Ma, B. Wang, J. Zhou, R. Jia, P. Luo, H. Wang, M.A. Mohamed, An effective risk identification method for power fence operation based on neighborhood correlation network and vector calculation, Energy Rep. 7 (2021) 6995–7003, https://doi.org/10.1016/j.egyr.2021.10.061.
[18] H. Wang, B. Wang, P. Luo, F. Ma, Y. Zhou, M.A. Mohamed, State evaluation based on feature identification of measurement data: for resilient power system, CSEE Journal of Power and Energy Systems 8 (4) (2021) 983–992, https://doi.org/10.17775/CSEEJPES.2021.01270.
[19] Y.Q. Huang, J.C. Zheng, S.D. Sun, et al., Optimized YOLOv3 algorithm and its application in traffic flow detections, Appl. Sci. 10 (9) (2020) 3079–3088, https://doi.org/10.3390/app10093079.
[20] Y. Tian, G. Yang, Z. Wang, et al., Apple detection during different growth stages in orchards using the improved YOLO-V3 model, Comput. Electron. Agric. 157 (2019) 417–426, https://doi.org/10.1016/j.compag.2019.01.012.
[21] G. Guo, N. Zhang, A survey on deep learning based face recognition, Comput. Vis. Image Understand. 189 (5) (2019) 102805–102816, https://doi.org/10.1016/j.cviu.2019.102805.
[22] X. Wang, S. Wang, J. Cao, et al., Data-driven based tiny-YOLOv3 method for front vehicle detection inducing SPP-net, IEEE Access 8 (7) (2020) 110227–110236, https://doi.org/10.1109/ACCESS.2020.3001279.
[23] J. Wang, J. Luo, B. Liu, et al., Automated diabetic retinopathy grading and lesion detection based on the modified R-FCN object-detection algorithm, IET Comput. Vis. 14 (1) (2020) 1–8, https://doi.org/10.1049/iet-cvi.2018.5508.
[24] C. Ding, D. Tao, Robust face recognition via multimodal deep face representation, IEEE Trans. Multimed. 17 (11) (2015) 2049–2058, https://doi.org/10.1109/TMM.2015.2477042.
[25] S. Jia, C. Hu, X. Li, et al., Face spoofing detection under super-realistic 3D wax face attacks, Pattern Recogn. Lett. 145 (7) (2021) 103–109, https://doi.org/10.1016/j.patrec.2021.01.021.
[26] W. Deng, L. Zheng, Y. Sun, et al., Rethinking triplet loss for domain adaptation, IEEE Trans. Circ. Syst. Video Technol. 31 (1) (2020) 29–37, https://doi.org/10.1109/TCSVT.2020.2968484.
[27] F.C. Stanford, N. Alfaris, G. Gomez, et al., The utility of weight loss medications after bariatric surgery for weight regain or inadequate weight loss: a multi-center study, Surg. Obes. Relat. Dis. 13 (3) (2017) 491–500, https://doi.org/10.1016/j.soard.2016.10.018.
[28] B. Xu, W. Wang, L. Guo, et al., CattleFaceNet: a cattle face identification approach based on RetinaFace and ArcFace loss, Comput. Electron. Agric. 193 (12) (2022) 106675–106688, https://doi.org/10.1016/j.compag.2021.106675.
[29] P.F. Borowski, Digitization, digital twins, blockchain, and industry 4.0 as elements of management process in enterprises in the energy sector, Energies 14 (7) (2021) 1885, https://doi.org/10.3390/en14071885.
[30] A. Di Tommaso, A. Betti, G. Fontanelli, et al., A multi-stage model based on YOLOv3 for defect detection in PV panels based on IR and visible imaging by unmanned aerial vehicle, Renew. Energy (Jun.) (2022) 193–201, https://doi.org/10.1016/j.renene.2022.04.046.
[31] D. Shi, H. Tang, A new multiface target detection algorithm for students in class based on bayesian optimized YOLOv3 model, Journal of Electrical and Computer Engineering 2022 (2022) 1–12, https://doi.org/10.1155/2022/4260543.
[32] Lei Feng, Bo Wang, Fuqi Ma, Hengrui Ma, Mohamed A. Mohamed, Identification of key links in electric power operation based-spatiotemporal mixing convolution neural network, Comput. Syst. Sci. Eng. 46 (2) (2023), https://doi.org/10.1155/2023/6870147.
[33] L. Wang, B. Wang, S. Wang, F. Ma, X. Dong, L. Yao, H. Ma, M.A. Mohamed, An effective method for sensing power safety distance based on monocular vision depth estimation, International Transactions on Electrical Energy Systems 2023 (2023), https://doi.org/10.1155/2023/8480342.
[34] Z. Chen, C. Ma, J. Ren, F. Hao, Z. Wang, Research on the identification method of safety wearing of electric power workers based on deep learning, Front. Energy Res. 10 (2023) 1091322, https://doi.org/10.3389/fenrg.2022.1091322.
[35] S. Subramaniam, R. Lakshmipathi, M.A.M. Arumugam, Evaluation of feature extraction and dimensionality reduction algorithms for face recognition using ORL database, in: Proceedings of the 2009 International Conference on Image Processing, Computer Vision, & Pattern Recognition, IPCV 2009, July 13-16, 2009, Las Vegas, Nevada, USA, 2 Volumes, DBLP, 2009, https://doi.org/10.1039/A708748I.
[36] S. Mitra, M. Savvides, Gaussian mixture models based on the phase spectra for illumination invariant face identification on the Yale database, in: 2007 FIRST IEEE INTERNATIONAL CONFERENCE ON BIOMETRICS: THEORY, APPLICATIONS AND SYSTEMS, WOS, 2007 000255193800030.