

Summary of Research on Prevention and Control Technology based on Deep Learning

Xiaoya Yang^{1,a}, Miaolei Deng¹, Hui Gao¹, Dexian Zhang¹

¹School of Information Science and Engineering, Henan University of Technology, Zhengzhou, Henan, 450001, China.

^aadmlei2003@163.com

Abstract

Under the background of anti-terrorism globalization, the prevention and control of violent terrorism poses new challenges to the field of artificial intelligence. This paper first summarizes the relevant knowledge of deep learning, then analyzes the crowd behavior analysis model based on deep learning, the detection and identification model of sensitive personnel and dangerous goods based on deep learning, the object behavior analysis and target tracking model based on deep learning, then suggested deficiencies and amendments. Finally, the anti-terrorism prevention and control work based on deep learning is summarized, the shortcomings are put forward, and the prospect is given.

Keywords

Deep Learning; Convolutional Neural Networks; Artificial Intelligence; Object Detection and Recognition; Behavior Analysis; Target Tracking.

1. Introduction

With the increasing frequency of global economic and cultural exchanges and the influence of extremism and terrorism, terrorist activities in sensitive areas are characterized by globalization -- from internal activities to external collusion, complexity and variety -- from unorganized to deliberate planning, regional transfer -- infiltration from remote areas to inland China, upgrading of means -- from cutting wounds and burns to large-scale explosive attacks, low-age crime -- from adult men to women and children, and goal generalization. In view of these characteristics, our country's anti-terrorism prevention and control work mainly focuses on technological innovation, judicial balance, legislative protection and so on. At present, the situation of prevention and control of violent terrorism in our country is still grim, and the task of resolutely combating terrorism will be completed for a long time with the joint efforts of all ethnic groups, regions and even all mankind.

Deep learning is the most important branch of machine learning in the field of artificial intelligence. Since the advent of deep learning, because of its strong learning ability, wide coverage and good portability, it has been favored by relevant researchers and has been well developed and applied in Computer Vision, Automatic Speech Recognition and Natural Language Processing. This paper focuses on its application in Computer Vision, analyzes the crowd behavior analysis model, object detection and recognition model, object behavior analysis and target tracking model based on deep learning. We are committed to the prevention and control of counter-terrorism, especially in sensitive areas of important meetings, events, rallies and demonstrations and infrastructure, and make use of in-depth learning technology to realize effective early warning and risk assessment of terrorist activities, to collect, analyze and share information in a timely manner, and to establish a modern public security anti-terrorism prevention and control system that is intensive, efficient, responsive, early warning, timely, safe and reliable.

2. Overview of Relevant Technologies

This section introduces the basic theory of deep learning involved in this paper. On the basic units of deep learning, such as convolutional neural network (CNN), generative adversarial network (GAN), long and short memory network (LSTM), this paper discusses the technology of prevention and control of terror based on deep learning. The historical block diagram of the development of deep learning is shown in figure 1.

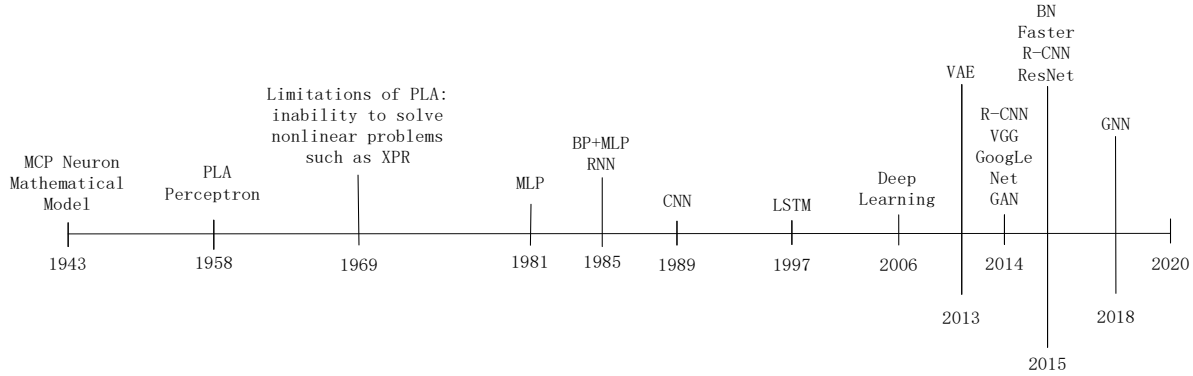


Figure 1. Development History of Deep Learning

2.1. Deep Learning and CNN

The framework of deep learning originated from early artificial neural networks. The early artificial neural network was influenced by biological neuroscience to simulate the information processing process of human brain, represented by MLP multi-layer perceptron [1], and then the progress of deep learning came from mathematical theory. Representative algorithm is back-propagation (BP) algorithm [2], since then depth learning model continues to improve.

$$\chi_j^l = f(\sum_i \chi_i^{l-1} * k_{ij}^l + b_j^l) \quad (1)$$

CNN is the most successful model in the deep learning framework. The convolution form is shown in formula 1. The core of the convolution kernel is the global sliding of the local perceptual convolution kernel in the upper map, and the sliding process shares the convolution kernel weight globally. With the help of multi-layer design of depth neural network, image spatial feature extraction is realized. Typical image recognition networks have AlexNet [3], VGG [4], ResNet [5] and so on, and task scenarios include detection, recognition, tracking and so on.

2.2. GAN

GAN network [6] was proposed in 2014. The core is to generate the parameter update of the network by game between the generator neural network and the discriminator neural network, which depends on the abstract evaluation of the discriminant network generation effect. In order to obtain a deep neural network model in line with the task target. GAN overall loss is designed as follows:

$$\min_G \max_D E_{x \sim p_r} [\log D(x)] + E_{x \sim p_g} [\log(1 - D(x))] \quad (2)$$

Since the traditional GAN two subnetworks have the disadvantages of training instability, gradient disappearance, pattern collapse and so on, the improved GAN include SRGAN [7], WGAN [8], MaskGAN [9] and so on. GAN network is also widely used in violence prevention and control because of its powerful abstract image feature perception ability.

2.3. RNN and LSTM

RNN is a kind of neural network that describes the characteristics of information timing, which is often used to process data such as text, speech and so on. The basic structural unit and time

line expansion of the RNN are shown in figure 2. In the standard RNN structure, the neurons of the hidden layer are weighted. As the timing advances, the front hidden layer will affect the latter hidden layer.

$$Q_t = g(V \cdot S_t) \quad (3)$$

$$S_t = f(U \cdot X_t + W \cdot S_{t-1}) \quad (4)$$

The S_t is the value of the hidden layer, and the output value is O_t , S_t depending not only on the X_t , but also on the S_{t-1} .

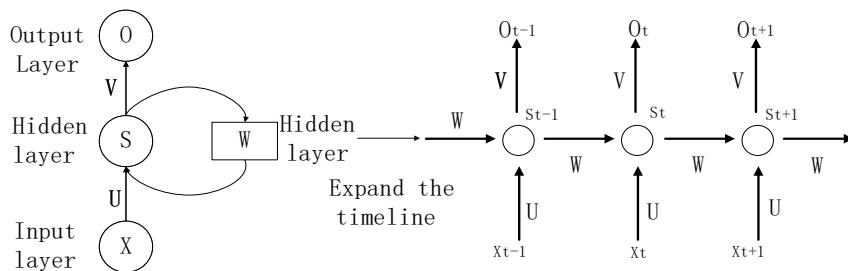


Figure 2. RNN Basic structural unit and timeline expansion

LSTM door structure is shown in Figure 3, Remember the values of different time lengths, LSTM different from the RNN, the repeat module has three multiplication operations to realize the forgetting gate, the input gate, the output gate, Record the discarded information, the collected information of the input, and the output information after processing respectively. To some extent, LSTM overcome the problems of RNN gradient vanishing, good performance in translation, emotional analysis, improved models include ON-LSTM [10], GRN [11], Deep RNN [12], etc.

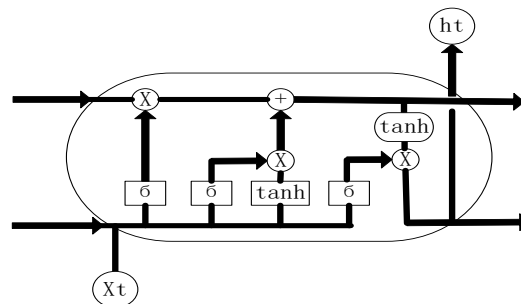


Figure 3. LSTM Door Structure

3. Prevention and Control Technology of Violent Crime Based on Deep Learning

Eliminating terrorist cases in swaddling clothes is the key to the prevention and control of violent terrorism. By analyzing the behavior of the crowd, the method of deep learning is used to effectively warn the sensitive personnel, dangerous equipment and violent terrorist behavior, and the abnormal target tracking model of sensitive persons and dangerous appliances is established to monitor the violent terrorist activities to realize the prevention and control of violent terrorist activities.

3.1. Population Behavior Analysis Based on Deep Learning

In the central city with large flow of people, the behavior analysis of the crowd is helpful to ensure the safety of public places, find out the hidden dangers in time, and avoid the occurrence of large-scale dangerous events or even violent terrorist cases. With the large-scale popularization and intelligent development of video surveillance equipment, we should make

full use of the effective information in the surveillance video to accurately locate and warn the abnormal events in time, and buy valuable time for the police personnel. Avoid further deterioration of violent terrorist incidents. Abnormal events are mainly different from the normal orderly flow of people disorderly chase, fight, sharp aggregation and diffusion.

The crowd behavior analysis algorithm is divided into four main directions: pedestrian detection, crowd count, crowd density analysis and crowd state analysis. Here we focus on the analysis of population density, usually using CNN to estimate the number of people, there are generally two schemes, one is to input the image, the number of people in the output area, the other is to input the population density map, integral to find the total number. To solve the general problem of crowd counting, Zhang [13] and others put forward two learning goals of crowd density and crowd counting, and use two CNN models to train alternately to keep a certain correlation between them and find out the local optimal solution. An adaptive CNN training model that does not require manual tagging is proposed when dealing with invisible target population scenarios, The MAE on UCSD and UCF CC 50 datasets are 1.60 and 467.0, MSE 3.31 and 498.5, it shows that the method is effective and reliable. To solve the multi-scale problem, Zhang [14] and others proposed a multi-branch convolutional neural network structure (MCNN), Each branch of the MCNN learns features that adapt to changes in target size so that it can estimate population size by population density, the validity of the MCNN model is verified on its own ShanghaiTech data set and all the data sets at that time. Feng [15] and others add video timing features to the analysis, propose a convolution LSTM (ConvLSTM) depth model and extend it to a bidirectional ConvLSTM model to obtain bidirectional features in time and space. After a series of transfer learning tests, it is proved that the applicability of the ConvLSTM is good.

The algorithms of crowd behavior analysis based on deep learning have emerged in recent years, Slicing CNN model [16], SaCNN model [17], ResnetCrowd model [18] have good performance in crowd behavior analysis. Current relative standard data sets have ShanghaiTech dataset, WorldExpo's dataset, UCF CC 50dataset, SmartCity dataset and so on. The focus of crowd behavior analysis in violent terrorist cases is the diversified data sets of scenes, the full use of space-time information of video images, multi-feature fusion technology, deep optimization of network structure and so on.

3.2. Object Detection and Recognition Technology Based on Deep Learning

Object detection and recognition is a common task in the field of Computer Vision. The detection objects in violent environment are generally divided into two categories: face recognition of sensitive personnel and detection of dangerous goods. The detection and identification of sensitive personnel in sensitive areas is conducive to the timely detection and investigation of persons who may cause terrorist cases, and the multiple detection and identification of dangerous goods. The credibility of terrorist incidents based on this is higher, which can greatly weaken the threat caused by terrorist cases.

3.2.1. Detection and Identification of Dangerous Goods Based on Deep Learning

Dangerous goods detection technology is most common in railway station and airport security inspection. At present, the main detection methods are X-Ray, Ion Mobility Spectrometry, Electron back scattered diffraction and so on. In the field of Deep Learning, Tian [19] changed the Fast RCNN VGG-16 network to 50 layers residual network (ResNet-50), replaced the Selective Search algorithm with Canny algorithm when generating the possibility region, the incremental network training is based on the original network training to maintain relevance and constraint. The key to the Faster RCNN lies in the mutual restriction balance between the Region Proposal Generating Network and the Target Classification with Bounding-box Regression Network in incremental training. On the Kitti data set, the Faster RCNN incremental network can not only detect the 5 types of vehicle targets of the original network, but also detect

the human targets of the new 3 actions. when the 11 types of security machine dangerous goods data set containing rifles, kitchen knives, swiss knives, combined knives as the new categories, the detection accuracy of some items is significantly improved, which verifies the effectiveness and reliability of the Faster RCNN incremental learning algorithm. For the detection of exposed and hidden dangerous goods, Yuan [20] proposed a depth-based approach Learning two-level detection identification method, two-level detection method corresponding to the treatment of exposed visible, hidden invisible dangerous goods identification. The first-level detection uses the improved GoogLeNet network to extract features and the BP neural network as the first-level classifier to jointly fine-tune the images classified as terahertz (THZ) categories. The THZ image is processed by range segmentation and image connected domain discriminant algorithm, and the second-level detection recognition is carried out by combining GoogLeNet and BP neural network. The natural image containing 8 kinds of samples of pistol and sniper gun is built by itself the data set 1 and terahertz images containing 6 samples of wrenches and axes demonstrate that the fine-tuned GoogLeNet outperforms VGG-16. Based on the YOLOV3 model, Xu [21] using k-fold method to preprocess the training data to solve the problem of overfitting, An improved Dense-YOLOv3 solves the gradient vanishing problem, soft-NMS algorithm is used instead of NMS algorithm to solve overlap and occlusion problems. Compared to the original YOLOv3 model, 9 classes of hazardous substances increased MAP by 9.65 per cent, Enhanced AP of objectives, Especially in the detection of knife, folding knife and other similar contour of small targets AP raised more than 15. In addition, the deep learning model used in dangerous goods detection also has the of MS-YOLOv3 [22], CNN THZ detection model [23].

In the work of violence prevention and control, the difficulty of the detection and identification technology of dangerous goods lies in the fact that the unified standardized dangerous goods data set has not been established, the optimization of deep network learning model, and the reasonable modeling of sensitive areas according to the demand of violence prevention and control.

3.2.2. Face Recognition Technology for Sensitive Personnel Based on Deep Learning

The detection and identification of sensitive personnel in sensitive areas is an important link in the technology of prevention and control of violence and terrorism, which plays a preliminary role in the timely attack of precision crime. Face recognition technology has been widely used in people's daily life, and face recognition technology based on depth learning has been relatively mature. When dealing with face recognition of video images, Zhao [24] put forward this multi-frame graphics fusion optimization model on the basis of GAN model, which has clear ROC performance advantages. When optimizing the bottom face quality, a twin GAN model based on cyclic consistency loss is designed. The face recognition model is introduced to provide ID feature supervision and pixel level MSE supervision respectively. Experiments show that the model improves the comprehensive performance of face recognition in violent terrorist scenes. Detection and recognition of low resolution face images, Ding [25] proposed multi-region convolutional neural network (MRCNN) model with sufficient training samples, The accuracy on CMU PIE and Extended Yale B data sets exceeds that of PCA LFDA, Bicube PCA (LFDA), "cross-space method", CCA and so on. Qin [26] combined with region segmentation with certain pose robustness Haar-SIFT depth belief network algorithm and SIFT sparse depth belief network algorithm with certain expression robustness, Optimizing sparse coding using local variance similarity, the recognition rates of 98.12% and 97.05% were obtained on FERET and AR face banks using deep belief network classifiers. In addition, The face recognition model based on depth learning also has IDICN model [27], DeepID 2 model [28], SBD2L model [29] and so on.

In the work of violence prevention and control, the difficulty of face recognition technology lies in eliminating the interference factors such as low resolution, occlusion, expression and so on,

improving robustness, data fusion and intelligence sharing of sensitive personnel in different sensitive areas, high-end equipment requirements for face data acquisition, and the construction of information research system dependent on recognition accuracy.

3.3. Object Behavior Analysis and Tracking Model Based on Deep Learning

In the prevention and control of violent terrorism, only the detection and recognition of dangerous goods and sensitive faces can only deal with some premeditated terrorist activities. In the face of sudden terrorist activities, real-time detection and judgment are often needed. Early warning attention to the behavior habits and trajectory of dangerous people can eliminate the threat of violent terrorist cases.

3.3.1. Object Behavior Analysis Based on Deep Learning

In the prevention and control of and control, it is not enough to rely on the detection and recognition of sensitive faces in known data sets to check suspicious molecules. We can analyze the information transmitted by the subject through body language to judge the rationality and danger of the subject movement.

The existing behavior analysis algorithms are mainly based on video series methods, Based on Two Stream Network [30] (TSN), Christoph [31] and others use CNN networks for spatial and temporal fusion, the recognition accuracy of spatial and temporal networks after replacing them with VGG-16 networks is 92.5% and 65.4% respectively. Wang [32] and others verified RGB plus optical flow plus warped optical flow combined input works best, To compare GoogLeNet, VGGNet-16 and BN-Inception networks, BN-Inception works best, Joint optimization of input, network architecture, training strategies, The recognition accuracy in UCF101 and HMDB51 is 94.2% and 69.4% respectively. Since then, the TSN model has experienced fusion improvement of partial weighted fusion [33], input feature graph timing reasoning [34] and so on. Song [35] and others proposed a temporal attention model, which uses two LSTM networks to aggravate the attention allocation joints for important frames and important content, The node has time variability. Du [36] and others use 3D convolution kernel and 3D Pooling to construct the C3D model to process the video frame directly, and the processing speed can reach 600 fps on the NVIDIA1080 graphics card. Besides, there are improved I3D [37], T-C3D [38], P3D [39], CDC [40], R-C3D [41] and other models.

3.3.2. Target Tracking Model Based on Deep Learning

In the work of violence prevention and control, timely target tracking directly affects the follow-up arrangement of violence prevention and control. Goal tracking model based on deep learning has entered people's vision since 2015. on the basis of CNN, RNN, LSTM, the goal tracking model is improved continuously. the real-time tracking model for sensitive personnel, dangerous appliances, sensitive vehicles and so on in the work of prevention and control of violent terrorism has gradually become a research hotspot.

The current target tracking technology is developing to pixel level tracking, multi-target tracking, 3D target tracking techniques for single target tracking and self-driving are relatively common. The MDNet [42] model based on deep learning wins the title in the VOT goal tracking competition in 2015. Each domain, is trained to update the parameters of the shared layer at the same time. The tracking stage uses the trained CNN network to realize the tracking location. C-COT [43] model won the 2016 VOT competition, which proposed continuous convolution operator operation and constructed continuous space learning problem with an implicit difference model. The second TCNN [44] model uses a multi-objective CNN network with tree structure to solve the problems of occlusion and target loss in tracking. Mueller [45] and others improved the traditional CF algorithm in 2017, and the proposed CA-CF algorithm added more background information for correlation filtering, and the tracking effect was obvious. Subsequently Sun [46] and others used nuclear ridge regression analysis (KRR) and CNN

complementary tracking, The former focuses on the overall information of the goal, The latter focuses on local information about the target, and finally combines the regression results, which complement each other to achieve accurate positioning of the object. Wang [47] and others divided the training process into two parts, the former tracking and the latter tracking, GT template frames and future frames are used to predict each other. In addition, The proposed Target-ware [48] mechanism, meta-learning [49] and MCCT [50] algorithms in recent years also show good results in the field of target tracking.

The challenges of target tracking include illumination, occlusion, plane rotation, motion blur, real-time and so on. The research of target tracking algorithm mainly focuses on correlation filtering and depth learning, while in the work of prevention and control of violent terrorism, multi-sensitive target tracking, target tracking in large-scale complex environment, background system sharing analysis of multi-video tracking information, intelligent prediction of integrated multi-target motion trajectory information and intelligent emergency strategy of possible events are all our future research directions.

4. Summary

This paper mainly focuses on the deep learning technology in the prevention and control of violent terrorism. This paper first summarizes the relevant knowledge of deep learning, then analyzes the crowd behavior analysis model based on deep learning, the detection and identification model of sensitive personnel and dangerous goods based on deep learning, the object behavior analysis and target tracking model based on deep learning, then suggested deficiencies and amendments. Finally, the anti-terrorism prevention and control work based on deep learning is summarized, the shortcomings are put forward, and the prospect is given.

Although terrorist activities at home and abroad have converged in recent years, the research on violence prevention and control technology based on deep learning technology has never stopped. In recent years, object detection and recognition technology based on deep learning is relatively mature, but behavior analysis and target tracking technology have not yet met the requirements of violence prevention and control. The future research focuses on object behavior analysis and target tracking. The work of prevention and control of violence and terrorism will develop towards the integrated public security system of active early warning and efficient study, data fusion and intelligence sharing, in-depth learning and technological innovation, and intelligent dispatching of anti-terrorism resources.

Acknowledgments

National Natural Science Foundation of China (2018YF*****02).

References

- [1] Rosenblatt F. PRINCIPLES OF NEURODYNAMICS. PERCEPTRONS AND THE THEORY OF BRAIN MECHANISMS [J]. American Journal of Psychology, 1963, 76(4).
- [2] Goh A T. Back-propagation neural networks for modeling complex systems[J]. Artificial Intelligence in Engineering, 1995, 9(3): 143-151.
- [3] Krizhevsky A, Sutskever I, Hinton G E, et al. ImageNet Classification with Deep Convolutional Neural Networks [C]. neural information processing systems, 2012: 1097-1105.
- [4] Simonyan K, Zisserman A. VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION[C]. computer vision and pattern recognition, 2014.
- [5] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[C]. computer vision and pattern recognition, 2016: 770-778.

- [6] Radford A, Metz L, Chintala S, et al. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks[J]. arXiv: Learning, 2015.
- [7] Wang X, Yu K, Wu S, et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks[J]. 2018.
- [8] Arjovsky M, Bottou L. Towards Principled Methods for Training Generative Adversarial Networks[J]. arXiv: Machine Learning, 2017.
- [9] Lee C, Liu Z, Wu L, et al. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation.[J]. arXiv: Computer Vision and Pattern Recognition, 2019.
- [10] Shen Y, Tan S, Sordoni A, et al. Ordered Neurons: Integrating Tree Structures into Recurrent Neural Networks [C]. international conference on learning representations, 2019.
- [11] Chung J, Gulcehre C, Cho K, et al. Empirical evaluation of gated recurrent neural networks on sequence modeling[J]. arXiv: Neural and Evolutionary Computing, 2014.
- [12] Gallicchio C. Short-term Memory of Deep RNN[J]. arXiv: Learning, 2018.
- [13] Zhang C, Li H, Wang X, et al. Cross-scene crowd counting via deep convolutional neural networks[C]. computer vision and pattern recognition, 2015: 833-841.
- [14] Zhang Y, Zhou D, Chen S, et al. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network[C]. computer vision and pattern recognition, 2016: 589-597.
- [15] Xiong F, Shi X, Yeung D, et al. Spatiotemporal Modeling for Crowd Counting in Videos[C]. international conference on computer vision, 2017: 5161-5169.
- [16] Shao J, Loy C C, Kang K, et al. Slicing Convolutional Neural Network for Crowd Video Understanding [C]. computer vision and pattern recognition, 2016: 5620-5628.
- [17] Zhang L, Shi M, Chen Q, et al. Crowd Counting via Scale-Adaptive Convolutional Neural Network[C]. workshop on applications of computer vision, 2018: 1113-1121.
- [18] Marsden M, McGuinness K, Little S, et al. ResnetCrowd: A residual deep learning architecture for crowd counting, violent behaviour detection and crowd density level classification[C]. advanced video and signal based surveillance, 2017: 1-7.
- [19] Tian Minhao. Research on Dangerous Goods Detection Algorithm of Security Inspection Machine Based on Incremental Learning [D]. North University of China, 2019.
- [20] Yuan Jianxin. Research on Dangerous Equipment Detection and Recognition Method Based on Deep Learning [D]. Dalian University of Technology, 2018.
- [21] Xu Chenyu. Research and improvement of dangerous goods detection model based on YOLOv3 [D]. Lanzhou University, 2019.
- [22] Zhang H, Xue B, Zhi Q, et al. Detection and Identification Method of Dangerous Goods under X-ray Based on Improved YOLOv3[C]. IEEE symposium series on computational intelligence, 2019.
- [23] Zhang Kaige, Gong Yaqiao, Zhao Guangzhou, etc. A Terahertz Dangerous Goods Detection Method Based on Deep Learning:.
- [24] Zhao Kai. Research on Key Technologies of Violent Terror Video Recognition Based on Deep Learning [D]. Beijing University of Posts and Telecommunications, 2019.
- [25] Chunhui Ding, Tianlong Bao, Saleem Karmoshi, Ming Zhu. Low-Resolution Face Recognition via Convolutional Neural Network[C]. IEEE Beijing Section, Guangdong University of Technology, University of Electronic Science and Technology of China. Proceedings of 2017 9th IEEE International Conference on Communication Software and Networks (ICCSN 2017). IEEE Beijing Section, Guangdong University of Technology, University of Electronic Science and Technology of China: IEEE BEIJING SECTION (Transnational Institute of Electrical and Electronics Engineers Beijing Branch), 2017:1187-1191.
- [26] Qin Qin. Research on face recognition based on improved SIFT deep belief network [D]. North China University of Technology, 2019.
- [27] Wu F, Jing X, Dong X, et al. Intraspectrum Discrimination and Interspectrum Correlation Analysis Deep Network for Multispectral Face Recognition [J]. IEEE Transactions on Systems, Man, and Cybernetics, 2020, 50(3): 1009-1022.

- [28] Sun Y, Chen Y, Wang X, et al. Deep Learning Face Representation by Joint Identification-Verification [C]. neural information processing systems, 2014: 1988-1996.
- [29] Chen Z, Wu X, Yin H, et al. Noise-robust dictionary learning with slack block-Diagonal structure for face recognition [J]. Pattern Recognition, 2020.
- [30] Simonyan K, Zisserman A. Two-Stream Convolutional Networks for Action Recognition in Videos[C]. neural information processing systems, 2014: 568-576.
- [31] Feichtenhofer C, Pinz A, Zisserman A, et al. Convolutional Two-Stream Network Fusion for Video Action Recognition[C]. computer vision and pattern recognition, 2016: 1933-1941.
- [32] Wang L, Xiong Y, Wang Z, et al. Temporal Segment Networks: Towards Good Practices for Deep Action Recognition [C]// ECCV 2016. Springer International Publishing, 2016.
- [33] Lan Z, Zhu Y, Hauptmann A G, et al. Deep Local Video Feature for Action Recognition[C]. computer vision and pattern recognition, 2017: 1219-1225.
- [34] Zhou B, Andonian A, Oliva A, et al. Temporal Relational Reasoning in Videos[C]. european conference on computer vision, 2018: 831-846.
- [35] Song S, Lan C, Xing J, et al. An End-to-End Spatio-Temporal Attention Model for Human Action Recognition from Skeleton Data[J]. arXiv: Computer Vision and Pattern Recognition, 2016.
- [36] Tran D, Bourdev L, Fergus R, et al. Learning Spatiotemporal Features with 3D Convolutional Networks [C]. international conference on computer vision, 2015: 4489-4497.
- [37] Carreira J, Zisserman A. Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset[C]. computer vision and pattern recognition, 2017: 4724-4733.
- [38] Kun L, Liu W, Gan C, et al. T-C3D: Temporal Convolutional 3D Network for Real-time Action Recognition [C]. national conference on artificial intelligence, 2018: 7138-7145.
- [39] Qiu Z, Yao T, Mei T, et al. Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks [C]. international conference on computer vision, 2017: 5534-5542.
- [40] Shou Z, Chan J, Zareian A, et al. CDC: Convolutional-De-Convolutional Networks for Precise Temporal Action Localization in Untrimmed Videos[C]. computer vision and pattern recognition, 2017: 1417-1426.
- [41] Xu H, Das A, Saenko K. R-C3D: Region Convolutional 3D Network for Temporal Activity Detection [C]// 2017 IEEE International Conference on Computer Vision (ICCV). IEEE, 2017.
- [42] Nam H, Han B. Learning Multi-domain Convolutional Neural Networks for Visual Tracking[C]. computer vision and pattern recognition, 2016: 4293-4302.
- [43] Danelljan M, Robinson A, Khan F S, et al. Beyond Correlation Filters: Learning Continuous Convolution Operators for Visual Tracking[C]. european conference on computer vision, 2016: 472-488.
- [44] Nam H, Baek M, Han B, et al. Modeling and Propagating CNNs in a Tree Structure for Visual Tracking[J]. arXiv: Computer Vision and Pattern Recognition, 2016.
- [45] Mueller M, Smith N, Ghanem B, et al. Context-Aware Correlation Filter Tracking[C]. computer vision and pattern recognition, 2017: 1387-1395.
- [46] Sun C, Wang D, Lu H, et al. Learning Spatial-Aware Regressions for Visual Tracking[C]. computer vision and pattern recognition, 2018: 8962-8970.
- [47] Wang N, Song Y, Ma C, et al. Unsupervised Deep Tracking[C]. computer vision and pattern recognition, 2019: 1308-1317.
- [48] Li X, Ma C, Wu B, et al. Target-Aware Deep Tracking[C]. computer vision and pattern recognition, 2019: 1369-1378.
- [49] Park E, Berg A C. Meta-tracker: Fast and Robust Online Adaptation for Visual Object Trackers[C]. european conference on computer vision, 2018: 587-604.
- [50] Danelljan M, Bhat G, Khan F S, et al. ATOM: Accurate Tracking by Overlap Maximization[C]. computer vision and pattern recognition, 2019: 4660-4669.