International Core Journal of Engineering Volume 9 Issue 9, 2023
ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0016

An Overview of the Application of Target Detection Technology in "Intelligent Construction Sites"

Zhihao Ni[1,2], Hong Song[3], Hao Wu[1,2]

[1] School of Automation and Information Engineering, Sichuan University of Science and Engineering, Yibin 644000, China

[2] Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Yibin 644000, China

[3] Aba Teachers University, Aba, Sichuan 623002, China

Abstract

With the continuous progress of science and technology, smart site technology plays an increasingly important role in the construction field. Target detection, as one of the core technologies, provides strong support for safety management, project efficiency, and resource utilization at smart construction sites. The aim of this paper is to synthesize the research progress and key technologies of target detection applications in smart construction sites, and to explore their potentials and challenges in improving site management and construction efficiency.

Keywords

Target Detection; Smart Construction Sites; Site Management.

1. Introduction

With the booming construction industry, construction safety is a growing concern[1-3]. Traditional human detection and sensor detection methods have certain limitations, with the development of machine vision technology, the detection of images and videos is becoming more and more mature, and the reduction of the corresponding accidents through intelligent detection technology is gradually becoming an important means [4-6]. "Intelligent construction site" [7] refers to a new type of project management method that utilizes advanced sensors, computer vision, artificial intelligence and other technological means to monitor and manage construction sites in real time. Target detection technology, as one of the core technologies in the "smart construction site", can identify and locate various targets in the construction site in real time, including personnel, vehicles, equipment, etc., and provide important information for site management and construction process.

2. Application of Target Detection in Smart Site

2.1 Security Management

On construction sites, target detection technology can help monitor safety within the site. By identifying dangerous behaviors and potential risks, such as people not wearing helmets, irregular operations, etc., timely warnings are issued and measures are taken to reduce the probability of accidents.

2.2 Engineering Efficiency Improvement

Target detection technology can be used to track construction progress and resource utilization on construction sites. By identifying and counting materials and equipment, the supply chain and

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0016

resource allocation can be adjusted in a timely manner to optimize the construction plan and improve construction efficiency.

2.3 Resource Management

On construction sites, target detection technology can also be used to manage and track all types of resources within the site. By identifying and monitoring equipment and materials within the construction site, comprehensive control of resources can be realized, reducing waste of resources and lowering construction costs.

3. Current Status of the Application of Target Detection Technology in Smart Construction Sites

3.1 Traditional Target Detection Methods

Traditional target detection methods are mainly based on image processing and computer vision techniques, such as Haar features, HOG features, and template matching based methods. Although these methods can realize target detection to a certain extent, they perform poorly for complex scenes and large variations of targets.

Traditional methods for target detection are generally divided into three stages: first selecting some candidate regions on a given image, then extracting features for these regions, and finally classifying them using a trained classifier. In the region extraction stage, the entire image is usually traversed using the sliding window method, while setting different scales and different aspect ratios to all targets. However, the computational cost of the sliding window method is relatively large, and the step size and scale requirements are high. Choosing a larger step size can reduce the number of input windows, thus reducing the computational amount, but less data will have an impact on the model performance; choosing a smaller step size will lead to too much input data and increased computational overhead. The main methods of feature selection are SIFT (Scale Invariant Feature Transform, SIFT) [8] and HOG (Histogram of Oriented Gradient, AQ, HOG) [9] . SIFT is a computer vision algorithm for detecting and characterizing localized features in images. The method extracts features by finding extreme points in the image and extracting position, scale and rotation invariants. It is worth noting that SIFT features are related to localized appearance points of interest on the object, independent of the image size and rotation, and have a high tolerance for light, noise, and slight viewpoint changes. HOG is composed by computing and counting histograms of gradient directions in localized regions of the image. Since HOG operates on localized square cells of the image, it maintains good invariance to geometric and optical deformations of the image, which only occur over larger spatial domains. Under coarse airspace sampling, fine orientation sampling, and strong local optical normalization, as long as the pedestrian maintains a largely upright posture, some subtle body movements can be tolerated, which can be ignored without affecting the detection results. However, HOG features are more difficult to deal with occlusion problems as well as larger object orientation changes. The main choices of classifiers are support vector machines (SVM) [10] and Adaboost [11]. SVM is a binary classification model whose basic model is a maximally spaced linear classifier defined on the feature space, a feature that distinguishes it from perceptual machines. Adaboost is an iterative algorithm whose core idea is to train different weak classifiers for the same training set, and then set the trained set of weak classifiers into a stronger final classifier.

In the existing work, these conventional algorithms are applied with helmet detection. For example, Rubaiyat [12] et al. first segmented the input image and then computed the discrete cosine transform to extract the HOG features from the discrete cosine transform, and then used Support Vector Machines (SVMs) to detect the presence of construction workers in the image. After detecting the construction workers a combination of color and Garden Hough Transform (CHT) based feature extraction techniques will be used for helmet detection. Feng Guochen [13] and others used Canny edge detection for human detection and then used color features to determine the location of the helmet, although the detection accuracy was high, but due to the small number of experiments, the phenomenon of missed detection was more serious.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0016

In summary, target detection based on traditional methods mainly suffers from the following problems: (1) the sliding window selection strategy is untargeted, the time complexity is high, and the window is redundant; (2) the robustness of manually designed features is poor, and the algorithm fails when the feature templates of the application scenarios are replaced.

3.2 Deep Learning Methods

In recent years, the rise of deep learning techniques has revolutionized target detection [14]. In particular, convolutional neural network (CNN)-based target detection methods, such as YOLO (You Only Look Once) [15], SSD (Single Shot MultiBox Detector) [16], and Faster R-CNN (Region-based Convolutional Neural Networks) [17], have greatly improved the accuracy and real-time performance of target detection. Traditional target detection traverses the features on the image by the sliding window method, which traverses the image by a fixed size and a fixed step size, although all the features can be extracted, the fixed window is unable to detect multi-scale objects, and at the same time there are more repetitive windows, which results in a lot of redundancy. In order to solve these problems, it is necessary to perform feature extraction on the image and then generate anchor frames on the feature map. Current algorithms in the target detection class can be categorized into two groups based on anchor frames: anchor frame based detection methods and anchor frame free detection methods.

3.2.1 Anchor Frame based Target Detection Algorithm

Anchor frame based target detection algorithms can be categorized into two groups: candidate region based algorithms (two-stage) and regression based target detection algorithms (one-stage). two-stage detection algorithms divide the detection problem into two stages, first generating candidate regions and then correcting and classifying the candidate regions. Typical of this type of algorithm is the candidate region based R-CNN [18] family of algorithms.

3.2.2 Target Detection Algorithm without Anchor Frames

Anchor frame-based detection algorithms need to artificially modify the size of the anchor frame for the dataset, resulting in poor generalization performance of the algorithm. To address this problem, the researchers proposed the anchorless frame detection algorithm. CenterNet [19] is improved and formed on the basis of CornrerNet, firstly, the input image is cropped to the size of 512*512, the image is inputted into the benchmark network to obtain a 128*128 Heatmap, the Box in the image is scaled to the Heatmap, the centroid coordinates of the Box are calculated, and the prediction box is determined according to the centroid coordinates by using Gaussian circle. Because the method can perform pixel-by-pixel level determination, it is widely used because it has some advantages over anchored frame networks in performing tiny target detection. Literature [20] firstly improves the problem of single decoding structure in CenterNet, combines the idea of spanning network such as U-net [30], and adds a long spanning layer in the same level of coding and decoding network, improves the fusion between the deep features and the shallow features, and reduces the loss of information in the process of convolution; and then joins the attention-boosting module and carries out the lightweight operation, which ensures the accuracy rate and at the same time meets the requirements of real-time detection on the site of the construction site. YOLOX [21] improved YOLOv3+DarkNet53 based on it using anchor-free detectors, simOTA, Multi positives, and other strategies to obtain high inference speed and detection accuracy. Literature [22] adds an attention mechanism to YOLOX and improves the loss function to enhance the network's detection accuracy for small targets while reducing the leakage rate for dense target detection.

4. Challenges Faced

(1) Construction sites are often in complex and changing outdoor environments, such as inclement weather and low-light backgrounds, which can affect the performance and accuracy of target detection algorithms. Pre-processing steps, such as rain and fog removal, low-light enhancement, etc., may be required for the application;

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0016

(2) The target detection algorithm needs to detect a large number of targets at the same time in practical applications, including workers, vehicles, equipment, etc., and even serious occlusion phenomenon, which puts forward higher requirements on the processing capability and real-time of the target detection algorithm.

(3) A large amount of video surveillance data involves workers' privacy, and how to protect the privacy and security of workers and related data is an urgent problem.

5. Future Outlook

With the continuous progress of artificial intelligence technology and the wide application of deep learning, target detection technology will continue to be improved and expanded. In the future, developments in the following directions can be foreseen:

(1) Introducing multimodal data fusion, combining image and LiDAR data, to improve the accuracy and robustness of target detection;

(2) Develop lightweight target detection models, such as mobilenet, ShuffleNet, Ghostnet, to reduce the number of parameters and decrease the computational complexity to improve the application of target detection on resource-constrained devices while ensuring the detection accuracy;

(3) Explore the application of target detection technology in areas other than smart construction sites, such as "smart cities".

6. Conclusion

Target detection, as an important technical means in "intelligent construction site", plays a vital role in improving the level of site management and construction efficiency. With the continuous progress of artificial intelligence technology, the prospect of target detection technology in the "intelligent construction site" is still full of hope. However, there are also a series of challenges to be faced, such as complex environments and data security, which require further research and exploration. Only on the basis of overcoming these challenges can target detection technology better bring more value to the smart site.

Acknowledgments

Sichuan University of Science and Engineering Graduate Student Innovation Fund (Y2022136).

References

[1] SIMONYAN K,ZISSERMAN ATwo-stream convolutional networks for action recognition in videos[EB/OLJ.(2014-11-12)[2022-04-13].https://arxiv.org/abs/1406.2199.

[2] FEICHTENHOFER C,PINZA,ZISSERMAN A.Convolutional two-stream network fusion for video actionrecognition[C]/2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Lasvegas,NV,USA:2016-06-27,2016:1933-1941.

[3] WANG LM.XIONG Y J,WANG Z,et al.Temporal segment networks:towards good practices for deep actionrecognition[C]/14th European Conference on Computer Vision(ECCV 2016),Amsterdam,The Nertherlands: 2016-10-11,2016:20-36.

[4] CARREIRAJ,ZISSERMAN A.Quo vadis,action recognition?a new model and the kinetics dataset[C]/2017IEEE Conference on Computer Vision and Pattern Recognition(CVPR),Honolulu,HI, USA:2017-07-21,2017:4724-4733.

[5] JI S, XU W,YANG M, et al. 3D convolutional neural net-works for human action recognition[J].IEEE Transactionson Pattern Analysis and Machine Intelligence, 2013,35(1):221-231.

[6] TRAN D, BOURDEV L, FERGUS R, et al. Learning spa-tiotemporal features with 3D convolutional networks[C]//Proceedings of the 15th IEEE International Conference on Computer Vision, Santiago,Dec 7-13, 2015. Piscataway:IEEE,2015:4489-4497.

[7] Kuenzel R, Teizer J, Mueller M, et al. SmartSite: Intelligent and autonomous environments, machinery, and processes to realize smart road construction projects[J]. Automation in Construction, 2016, 71: 21-33.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202309_9(9).0016

[8] LOWE D G.Distinctive image features from scaleinvariant keypoints[J].Intemnational Journal of Computervision, 2004,60(2):91-110.

[9] FEICHTENHOFER C,PINZA,ZISSERMAN A.Convolutional two-stream network fusion for video actionrecognition[C]/2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Lasvegas,NV,USA:2016-06-27,2016:1933-1941.

[10] ABDOLLAHI S, POURGHASEMI H R, GHANBARIAN G A, et al. Prioritization of effective factors in the occurrence of land subsidence and its susceptibility mapping using an SVM model and their different kernel functions[J]. Bulletin of Engineering Geology and the Environment, 2019, 78(6): 4017–4034.

[11] SIMONYAN K,ZISSERMAN ATwo-stlream convolutional networks for action recognition in videos[EB/OLJ.(2014-11-12)[2022-04-13].https://arxiv.org/abs/1406.2199.

[12] Rulbsiyat A H M, Toma T T,KalantariKhandani M,et al,Automatic Dxtecxtion ol Helmet UIses lfor Construction Sulety[C]//2016 IEEE/WIC/ACM International Conference on Web Intelligcnce Workshop. Omaha; IEEE,2016;135-142.

[13] FENG Guochen, CHEN YanyanResearch on automatic identification technology of the safety helmet based on machine vision[J]. Machine Design and Manufacturing Engineering,2015,44(10);39 42.

[14] DALALN,TRIGGS B.Histograms of oriented gradients for human detection[C]//2005 IEEE Computer SocietyConference on Computer Vision and Pattern Recognition(CVPR'05),San Diego,USA,June 20-25,2005:886-893.

[15] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[16] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[17] Ren S,He K.Girshick R,et al.Faster R-CNN:Towards real-time object detection with region proposalnetworks[CAdvances in Neural Information Processing Systems,2015:91-99.

[18] Girshick R,Donahue J,Darrell T,et al.Rich feature hierarchies for accurate object detection and semanticsegmentation [CJIProceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014:580-587.

[19] Duan K, Bai S, Xie L, et al. Centernet: Keypoint triplets for object detection[C]//Proceedings of the IEEE/CVF international conference on computer vision. 2019: 6569-6578.

[20] SUN L,JIAK,YEUNG D, et al. Human action recognitionusing factorized spatio-temporal convolutional networks[Cv/Proceedings of the 15th IEEE International Conference onComputer Vision, Santiago, Dec 7-13, 2015. Piscataway:IEEE, 2015:4597-4605.

[21] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015: 234-241.

[22] Ge Z, Liu S, Wang F, et al. Yolox: Exceeding yolo series in 2021[J]. arXiv preprint arXiv:2107.08430, 2021.