JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

A Survey on Detection of Objects in Video Using Deep Learning

Ms Yasmeen K. Mulla*, Prof. K. T. Mane**

**M.Tech Student, Department of computer science and Engineering, D.Y Patil College of Engineering & Technology, Kolhapur*

***Assistant Professor, Department of computer science and Engineering, D.Y Patil College of Engineering & Technology, Kolhapur*

Abstract: Due to object detection is having a very close association with video analysis and understanding of different images, it has captured much attention. Existing object detection methods are based on features of images that are handcrafted and always show poor performance in the case of shallow training objects. It also not suitable for the detection of comparatively dense and small objects. Sometimes they unable to identify objects with any given geometric transformations. Their detection performance can easily show a stagnate while comparing to the high-level deeper features with the features of images of low-level In particular for the relevant circumstance, the standard surveillance system does not provide the object identification capability. The target can readily shift in diverse backgrounds with varying brightness and shadow in the conventional video object detecting system. In the case of complicated object detection, the conventional object detection system is unable to effectively identify the targets, especially in real-time. In this paper the different techniques for detection of objects are reviewed and based on the review different drawbacks and challenges has been identified to solve the problems of existing systems with respect to efficiency.

Keywords: Deep learning, key frames, Object Detection, LBPH, Eigen Faces, Fisher Faces

1. Introduction:

The need to automatically extract useful information from the video system is growing along with the widespread adoption of video surveillance systems. Target detection and behaviour identification are two areas where feature extraction performs well. Since deep learning makes use of local perception, weight sharing, and down sampling, it offers excellent tolerance against displacement, scaling, and distortion. Convolution neural networks (CNNs) are currently a powerful picture identification method that may be applied in a number of situations, such as face and licence plate recognition.

Face recognition is an important area of research for image analysis, in addition to many other applications used for essential aspects of effective communications, such as information security (application security, database security, file encryption, internet access, and medical records), law enforcement and surveillance (video surveillance, CCTV control), bioinformatics (driving, licence, passport, voter registration, aadhar card, etc.), smart cards, and access control. Similar to how they are drawn to pattern recognition, neural networks, computer vision, and graphics, many academics are interested in the idea of human face detection via deep learning.

The facial recognition system typically operates in two modes:

1. Facial Verification: This involves a one-to-one comparison of a query face image with database image.
2. Face Identification: This process entails comparing a query face image to all the frames kept in the database in a one-to-many fashion. The photos that are collected are pre-processed in order to obtain good recognition in any situation. As noise affects the majority of photos, one of these is noise removal.

2. Literature Review:

It has been acknowledged in Wenming Cao, et. al. [1] larger and deeper neural networks are steadily advancing the performance of various visual and machine learning tasks. They usually demand vast sets of labelled knowledge for good results and suffer from extraordinarily high procedural complexity as a result, making them frequently impossible to be deployed quickly for systems like vehicle object identification from vehicle cameras for power-assisted driving. This paper investigates the development of a quick deep neural network for real-time video object detection by studying the concepts of knowledge-guided coaching and predicted regions of interest. It specifically discusses the creation of a brand-new architecture for training deep neural networks on datasets with few labelled samples by utilising ready-to-use cross-network data projection.

To reduce the complexity of processing, the regions of interest are estimated using a mathematical formula. The experimental results on car detection from videos conclusively demonstrate that the suggested methodology is capable of up to sixteen times network acceleration while maintaining the article detection performance. The topic's primary focus is on the area of interest. Vehicle recognition is 1.6 times faster when the region of interest is immediately identified and a low-quality SVM model is used, yet accuracy is essentially same. It will perform sixteen times faster overall using the faster deep neural network, producing predictions from feature maps of various scales.

The final detection performance depends heavily on the estimation of salient regions. The vehicle's poor cantering at the calculable zone, which results in accuracy loss, will have an impact on the detection accuracy. A method to enhance the salient region estimation is also necessary. The optical flow suggested by Srivatsa

Prativadibhayankaram, et al.[2], is used in a compressive online strong Principal element analysis (RPCA) to iteratively separate a sequence of video frames into the foreground (sparse) and background (low-rank) parts. This separation methodology will approach each video frame from a limited low set of

measurements as opposed to batch-based RPCA, which normally processes all the data. Additionally, optical flow is used to estimate and then compensate for motions between prior foreground frames.

Several vision-based applications struggle with the problem of detecting moving objects in a lengthy video sequence, according to Lee, et. al.[3] 's research. moving items, notably in detective work. Once the camera is moving, it is difficult to identify. In this research, the author put forth a parallel technique for object movement while working with a dynamic background. The projected region of a specific object is first sighted using a backdrop compensation technique. He then introduced a lightweight convolutional neural network-based method for small objects called YOLOv3-SOD for detective work all things within the image, which is specifically made to discover moving objects. In the end, the results of the object detection and the moving objects area unit are combined.

The author of Berjón, D., et al. [4] developed a strategy to improve performance with huge, complex datasets. which methodology was determined to be the most complicated and time-consuming. Additionally, Bianco S, et al. [5] think Many computer viewing applications, such as video surveillance, smart locations, video identification, and retrieval, call for the finding of changes within the video streams. The solid switch algorithms have a low false alarm value, which is necessary as a pre-processing step in all of these applications. There were numerous algorithms put forth to address the issue of getting a video change. To distinguish the scene from the foreground and background parts, the majority of us rely on retrieval strategies. The approach for identifying binary images of the front areas corresponding to moving objects produces frequently changing results.

These algorithms are made to deal with foreseeable problems. In real-world videos, there are many natural variances, lighting changes, shadow changes, camera motions, cameracaused distortions, and so on. Algorithms have gotten more complex throughout time, which has made them more time and memory intensive. To speed up calculations for use in real-time applications, output back techniques could be aligned to the GPU.

The genetics-based evolution technique was first developed by J. Redmon et al. [7] and combines skills to discover a little alteration to produce a more effective algorithm. We can choose from the smaller set of straightforward algorithms thanks to the solutions offered by genetic programming. The integration technique can result in algorithms that resolve detect-changes issues more effectively. The learning framework has been proposed to facilitate the training of more in-depth networks than traditionally used [8]. The layers are classified as residual learning activities in terms of layer input, rather than reading nontargeted functions. The author provides ample evidence that these remaining networks are easy to use, and can gain accuracy from the most extensive depths. Lack of Trained data affects accuracy and confidence level. Improvisation is required of efficient training. Also Rios-Cabrera R,et.al. [16], The analysis is performed to determine which method has the highest accuracy of prediction rating for various test features. By supplying several sample images of different people acting as test subjects. The machine begins training to produce individual features between test subjects. After that, a new image was tested against the “educated” data and labelled as one of the subjects. By analysing the experimental data, the Eigenfaces method was determined to have the highest predictive rate of the three

tested algorithms. The author of paper [17] proposed a novel method for locating numerous 3D objects in daily life. starting with a method based on images. He advised us to read the illustrations with discrimination at the beginning. demonstrates that this may be done instantly and online during the gathering of test photos, greatly affecting the detector's accuracy. Second, he advises using a cascades-based algorithm to speed up detection. The measurement method is more efficient since new items may be added more quickly and for relatively little money, but the findings lack precision.

Pre-processing of the video is necessary to improve the detection of moving objects. The noise is in the form of disturbances which present in the background. The key frames are separated which obtained are free from unwanted noises, is used further for feature extraction The frames are separated called the key frames which are used for the matching. Further the features are extracted. The process of feature extraction is crucial to picture categorization. It enables the most accurate representation of visual material. The process of feature extraction involves removing from key frames the essential features needed for training and storing them separately. At first it is necessary to define minimum window size to be recognized as a face. For that height and width of rectangle is defined in advance so that each face will be easily recognizable. It mainly includes Edges, Corners, interest points, regions of interest. Then the training is performed. Using the data we have collected, this will be used to train the model. After training, the model will predict (based on confidence) if the user is the same user or someone else. We must first extract the data from the output file and convert it to grayscale before we can start the training. We then append the data to train data after this colour transition. Image classification is completed afterwards. The efficiency and precision of pattern recognition depend heavily on image classification, which is a required first step. A deep learning job called classification may identify the various items in a picture or video. To identify which items are present, it relates to training machine learning models. Classification is useful at the yes-no level of deciding whether an image contains an object or not.

Conclusion:

We have reviewed a large number of papers on video object detection. While much has been done, there are still many research opportunities in automatic video object detection. As most of the existing framework's lack in accuracy. The limited labelled data samples give the poor results causes loss of accuracy. Salient region detection (region of interest) becomes very crucial as data sample size becomes exponential. It effects on speed and performance. The current framework shows the lack of power in case of noise cancellation it performs better in limited datasets but loss speed in complex dataset. Although time is most crucial factor but noise cancellation, training and classification of dataset takes significant time. Time reduction is the prime requirement to boost the overall performance of the system.

The proposed system gives excellent performance in case of accuracy, as the results are been tested against the two other algorithms i.e., eigen faces and fisher faces. The LBPH algorithm works great with large data samples and even complex data sets. It's been less affected by noise and remarkable performance in limited time duration.

3. References:

[1] Cao, Wenming et al. "Fast Deep Neural Networks With Knowledge Guided Training and Predicted Regions of Interests for Real-Time Video Object Detection." IEEE Access 6 (2018): 8990-8999.

[2] Prativadibhayankaram, S.; Luong, H.V.; Le, T.H.; Kaup, A. Compressive Online Video Background–Foreground Separation Using Multiple Prior Information and Optical Flow. J. Imaging 2018, 4, 90.

[3] Lee, Sang-ha et al. "WisenetMD: Motion Detection Using Dynamic Background Region Analysis." Symmetry 11 (2019): 621.

[4] Berjón, Daniel et al. "Real-time nonparametric background subtraction with trackingbased foreground update." Pattern Recognit. 74 (2018): 156-170

[5] Bianco, Simone et al. "Combination of Video Change Detection Algorithms by Genetic Programming." IEEE Transactions on Evolutionary Computation 21 (2017): 914-928..

[6] Cioppa, Anthony & Droogenbroeck, Marc & Braham, Marc. (2020). Real-Time Semantic Background Subtraction.

[7] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[8] He, Kaiming et al. "Deep Residual Learning for Image Recognition." 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016): 770-778.

[9] Hoffman, Judy et al. "Simultaneous Deep Transfer Across Domains and Tasks." 2015 IEEE International Conference on Computer Vision (ICCV) (2015): 4068-4076. [10] Rios-Cabrera R, Tuytelaars T. Boosting masked dominant orientation templates for efficient object detection[J]. Computer Vision and Image Understanding, 2014, 120: 103-116

[10] A Survey of Deep Learning-based Object Detection Licheng Jiao, Fellow, IEEE, Fan Zhang, Fang Liu, Senior Member, IEEE, Shuyuan Yang, Senior Member, IEEE, Lingling Li, Member, IEEE, Zhixi Feng, Member, IEEE, and Rong Qu, Senior Member, IEEE

[11] Ajeet Ram Pathak, Manjusha Pandey, Siddharth Rautaray, Application of Deep Learning for Object Detection, Procedia Computer Science,Volume 132,2018, Pages 1706 1717, ISSN 1877-0509,

[12] R. Kalsotra and S. Arora, "A Comprehensive Survey of Video Datasets for Background ubtraction," in IEEE Access, vol. 7, pp. 59143-59171, 2019, doi 10.1109/ACCESS.2019.2914961.

[13] Mohana and HV Ravish Aradhya, "Object Detection and Tracking using Deep Learning and Artificial Intelligence for Video Surveillance Applications" International Journal of Advanced Computer Science and Applications(IJACSA), 10 (12),2019. http://dx.doi.org/10.14569/IJACSA.2019.0101269

[14] Sierra, Brandon Luis. "Comparing And Improving Facial Recognition Method."(2017).

[15] Rios-Cabrera, Reyes and Tinne Tuytelaars. "Discriminatively Trained Templates for 3D Object Detection: A Real Time Scalable Approach." 2013 IEEE International Conference on Computer Vision (2013): 2048-2055.

[16] Research on Face Recognition Classification Based on Improved GoogleNet Zhigang Yu, Yunyun Dong, Jihong Cheng, Miaomiao Sun , and Feng Su :January 2022

[17] Podgorsak AR, Rava RA, Shiraz Bhurwani MM, Chandra AR, Davies JM, Siddiqui AH, Ionita CN. Automatic radiomic feature extraction using deep learning for angiographic parametric imaging of intracranial aneurysms. J Neurointerv Surg. 2020 Apr;12(4):417-421. doi: 10.1136/neurintsurg-2019-015214

[18] Medjahed, Seyyid Ahmed. (2015). A Comparative Study of Feature Extraction Methods in Images Classification. International Journal of Image, Graphics and Signal Processing. 7. 16-23. 10.5815/ijigsp.2015.03.03.

[19] Deng, Jun & Xuan, Xiaojing & Wang, Weifeng & Li, Zhao & Yao, Hanwen &Wang, Zhiqiang. (2020). A review of research on object detection based on deep learning. Journal of Physics: Conference Series. 1684. 012028. 10.1088/1742-6596/1684/1/012028.

[20] Xin, Mingyuan & Wang, Yong. (2019). Research on image classification model based on deep convolution neural network. EURASIP Journal on Image and Video Processing. 2019. 10.1186/s13640-019-0417-8.