International Journal of Innovative Research in Computer Science & Technology (IJIRCST)
ISSN: 2347-5552, Volume-8, Issue-2, March 2020
https://doi.org/10.21276/ijircst.2020.8.2.2
www.ijircst.org

Detection of Human Behaviour by Object Recognition using Deep Learning: A Review

Mr. Utsab Mukherjee, Prof. Samir Kumar Bandyopadhyay

ABSTRACT- The major drawback of the society is the falsehood of human nature; so prediction of nature of the individual by analysing the video or image of that person is highly necessary. From the post World War II period due to the advancement of technology since the past few decades many countries have developed low cost cameras with high resolution. They generally use the RGB and depth features to enhance the image quality captured by the camera. Hence object recognition is the new branch of computer science which emerges. It has got a close relationship with image understanding and analysis of video; which encourages several researchers to work in this domain since past few years. As the branch of deep learning develops handful of efficient tools have been develop which prove to be highly efficient to learn high level deeper features of the image and semantics. In this paper a review on object recognition using deep learning has been discussed. This paper includes basics of deep learning and object recognition. Then we have discussed some tools required to perform object recognition as well as moving object recognition and finally some future goals of this subject have been discussed.

KEYWORDS: Deep Learning, Object Detection, Neural Network and Object Recognition

I. INTRODUCTION

The major problem of modern day society is the falsehood of human nature. It is evident that while interrogating any convicted person or criminal, the person does not tell the accurate truth or tends to hide something. To minimise the error of crime investigation we wish to incorporate computer vision into it. One of the promising fields in modern day is computer vision. The basic tool to perform this is object recognition [1]. Object Recognition has got a lot of significance in determining the identity of an object. This has a huge application in face recognition, gender recognition, as well as the far reaching goal is to detect the psychology of any individual. The objective of this research is to analyse the behaviour of any individual based on the person's image.

Manuscript received March 20, 2020

Mr. Utsab Mukherjee, Assistant Professor, Department of Computer Science, Bhawanipur Education Society College, West Bengal, India.

Prof. Samir Kumar Bandyopadhyay, Academic Advisor, Bhawanipur Education Society College, West Bengal, India. (email: 1954samir@gmail.com)

The accuracy can be enhanced if we do the same using moving object detection. Before going to the detailed study, let us state the definitions of deep learning and object detection.

Deep learning [3] is a branch of machine learning that has a significant network to learn unsupervised data that is not labelled. Human brain has got thousands of neurones by which we can identify any object. Deep learning is a method in which a network of such artificial neurones is used to identify any object with good accuracy.

In order to achieve overall knowledge of any particular image, apart from classifying or clustering we should be able to detect particular objects in an image; as well as identify the background or subject of the image. For this reason a new branch of computer science emerges known as object recognition [2]. Object recognition is the branch of study that takes from the real world and extract the object from it. Human beings use to see different objects and the brain with the help of thousands of neurones the particular object gets recognized with great accuracy. The purpose of this subject is to manipulate the algorithms performed by human beings to recognise objects. There are many applications of object recognition as well as moving object recognition: they are face recognition, skeletal recognition, autonomous driving, human nature analysis and many more. As one of the significant applications of computer vision, object recognition serves as the important tool to perform that.

The object recognition [2] problem can be defined as a classification problem based on models of identified things. Formally, given an image that contains one or more objects of interest (and background) and a set of labels corresponding to a set of models known to the system.

Object recognition or moving object detection can be done by using machine learning algorithms. Depending on the applications we can use supervised or semi-supervised learning. However, the object detection based on deep learning is an important application in deep learning technology, which is characterized by its strong capability of feature learning and feature representation compared with the traditional object detection methods. We prefer deep learning mechanism over supervised or semi supervised learning because the efficiency is better; apart from that deep learning takes much less time to train the known dataset.

II. PROCESS OF OBJECT RECOGNITION

The traditional object recognition is done in three major steps; they are: informative region selection, feature extraction and classification.

A. Informative Region Selection

Since different objects may emerge in random positions of the image having different aspect ratios or sizes, hence we scan the entire image with a multi-scale sliding window. Although this exhaustive method can figure out all possible positions of the objects, still it has certain drawbacks. Because of a significant number of candidate windows, it makes high computation cost producing many redundant windows. However, if only a constant number of sliding window templates are applied, unsatisfactory regions may be produced.

B. Feature Extraction

In order to recognize multiple objects, we must extract visual features; that provides a semantic and robust representation. SIFT[4], HOG[5] and Haar[6] like features are used. These features work almost analogous to the neurones of human brain.

C. Classification

A classifier is used to distinguish the extracted features from the image. This is used to detect the target object from the image. There are many available classifiers each used to serve a specific purpose. Usually, we use the Supported Vector Machine (SVM)[7] , AdaBoost[8] and Deformable Part-based Model (DPM)[9]. The application domain of object recognition is shown in Figure 1.

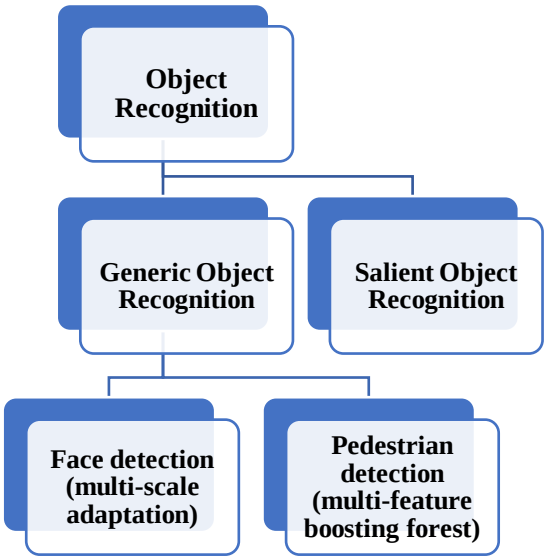

Figure 1: Application Domains of Object Recognition

III. LEARNING APPROACHES

In order to learn complex features some models are used having complicated architectures. Deep Neural Network(DNN)[10] makes a significant gain by introducing Regions with CNN features(R-CNN)[11]. DNN or CNN behaves quite differently with respect to traditional approaches. With the proposal of R-CNN, improved models have been recommended where fast RCNN optimises classification and regression tasks. YOLO can be used to perform object detection by fixed grid detection[12].

IV. RELATED WORKS

In this domain a lot of works have already been done. Some of them are discussed.

Mandal M, et al.[13] has presented a novel deep learning framework which performs online moving object detection while streaming videos. Pixel-wise binary segmentations were previously done whereas the OR technique does not distinguish between moving or static objects. At first Mandal M[13] has attempted to classify moving parts of the video and they have achieved this by labelling axis aligned bounding boxes for moving objects which requires less computational resources than producing pixel-level estimates. In their proposed MotionRec framework[13], they have learned both temporal and spatial features using past history and current frames respectively. Initially the authors have estimated the background with a temporal depth reductionist (TDR) block. Followed by the estimated background, current frame and temporal median of recent observations were assimilated to encode spatiotemporal motion saliency. From the motion saliency maps feature pyramids have been developed which performs regression and classification of fetched data. This framework[13] works online which requires few frames. The authors have 42614 objects in 24923 video frames.

Murlidharan R, et al[14] makes an introduction how deep learning is used in Object Detection, the authors have used SVM-KNN based model to perform object recognition. The authors have introduced classical methods in object recognition followed by deep learning algorithm has been implemented in their paper[14]. Their paper focuses on the framework design and working principle of the models then analyses the level of accuracy. The authors have proven that SVM and KNN mixed framework produces better efficiency for object recognition.

Alexandre L.A.[15] proposes a current trend in image processing data by using CNN. The his paper[15] he has investigated the possibility of transferring information between CNN while processing RGB-D data to improve the accuracy by reducing time requirement for training the data set. He has presented the experimental observation to show the above mentioned goal. He has proposed the use of four independently trained CNN one for each channel. He came to the conclusion that by using each CNN for all channels reduces the time for training data set as well as reduces errors. Alexandre[15] has also proposed the use of transfer learning between each CNNs for respective channels. This showed better result but the training time got increased. Hence he has proven that it is advantageous to split RGB-D data into four separate training sets.

Felzenszwalb P.F. et al[16] has described an object recognition model that is based on many multi-scale part models that are deformable. Their system is able to

represent object classes with large variables and achieved high level of accuracy. They have used pascal datasets and used partially labeled data. In their research work[16] they have combined a sensitive approach of data mining by using latent SVM. The training algorithm is iterative which alternates between fixing latent values for positive examples and optimises the SVM function. The authors wished to make grammar based models to represent objects with hierarchical structures. The future models should allow for mixture models as part level and should have reusability of parts.

Yang B, et.al.[16] proposed a model which can detect face efficiently as well as effectively. The authors have used the concept of channel features and incorporated the idea into face detection domain. This extends the image channel to varied types like oriented gradient histograms or gradient magnitudes which eventually encodes essential information in simpler forms. Hence the authors have adopted a novel variant known as 'aggregate channel features' which designs the feature entirely. This even enhances the multi-scale version of features with better result. Therefore a multi-view detection approach was adopted that features score re-ranking and detection adjustments. The researchers have used their algorithms on AFW and FDDB test sets which they have run at 42FPS in VGA images.

Chen X, et al.[17] proposed a model which has an aim to generate high quality 3D object recognition for autonomous driving. Their method[16] uses stereo imagery to place proposals in form of 3D bounded boxes. They have formulated the problem to detect depth patterns. Their experimental observation reviled good performance over traditional models. The efficiency further improves if CNN and the object proposal method works simultaneously.

Kang K, et. al.[18] has proposed a model for recognition of objects from videos by which consist of novel Tubelet Proposal Network which generates spatiotemporal proposals efficiently. In order to fully utilize the temporal information the state of the art methods that are based on spatiotemporal tubelets were used that are basically sequences of linked bounding boxes. An LTSM network has been incorporated with temporal information from tubelet proposals that achieves better accuracy in videos. The authors have done experiments on large scale ImageNet VID data set.

Paisitkriangkri S et al.[19] proposes a model which handles the high cost of training any classifier by learning suitable features of pedestrian detection. The authors have presented deep neural network for pedestrian recognition. They have used the learning process highly efficient because of using DNN. Furthermore a KNN method has been implemented to solve the comparison between region of interest and templates. The performance of their framework gets better for different data sets over traditional processes since the model has less dependency on the classifier. Their model[19] also works well in public datasets.

Byeon W et al[20] viewed the drawback of pixel-level classification and segmentation of images using LTSM learning approach. In their work, it has been investigated by the authors that 2D LTSM networks for natural scene images were labelled according to complex spatial dependencies. Previously all the other frameworks used separate modules for classification and segmentation of the image as well as poor post processing of the image; but in their method; classification, segmentation, and context integration were all carried out by 2D LSTM network that allows learning of texture and spatial parameters in a single module. It has been observed that their model[20] was capable to capture local or global contextual information over raw RGB values with good efficiency. The model was so designed that it had lower computational complexities with respect to previously used models. They have used SIFT Flow datasets. Since no pre or post processing is needed the task can be done by just a single core processor making the framework acceptable by the mass.

V. APPLICATIONS

There are many applications of object recognition; however we have studied 3D object recognition from any digital image and moving object detection. These two applications basically serve the purpose of behaviour analysis of any individual.

A. 3D object detection

With the advancement of technology the 3D sensors like LIDAR and digital cameras are developed which provide additional depth information apart from RGB data. This can be used for the understanding of 3D object in 2D image. This can be done by CNN[15] . These 3D aware techniques; places correct 3D bounding boxes around detected objects. Multi-view representation[16] or 3D proposal networks[17] can be done to encode depth information with necessary hardware.

B. Video object detection

One of the most promising fields of computer vision is moving object detection. For this motion rec[13] is a model that serves as one of the best framework for the recognition of objects from video. Here temporal information through multiple frames serves as the key role to understand the features of different objects. The major hardship that the researchers face is the accuracy varies if the video gets blurred or defocused. Spatiotemporal tubelets[18], optical flow[19] and LSTM[20] are some basic models by which video object detection are done. It is noted that pedestrian detection is also significant in traffic signal processing [19].

VI. FUTURE WORKS

Object recognition or moving object recognition has a great significance in the field of computer science. By this process many applications can be achieved. First, we use the model for human psychology prediction. Secondly, this model can play a significant role in monitoring unmanned traffic system. Thirdly, we can use

the concept of moving object detection in many online video streaming apps in which the review or rating can be analysed by simply detecting the viewer's facial expression. Most importantly, this can be implemented to detect driver's drowsiness by capturing real-time video by any camera placed on the windscreen of the car while he/she drives in a highway for hours.

VII. CONCLUSIONS

In order to detect any individual's behaviour by analysing him/her from any image or video we need object recognition or moving object recognition respectively. Object recognition can be done by many ways. Deep learning based object recognition in computer vision has becoming a topic with great interest among the researchers since it has powerful learning ability and advantageous in terms of dealing with scale transformation or background switches. In this paper we provide a review on object recognition using deep learning for different existing frameworks capable to handle different problems. We state the steps of object recognition and also state why deep learning is favourable over traditional learning methods. Nevertheless, in this paper we propose some promising future goals of the research.

ACKNOWLEDGEMENT

The Bhawanipur Education Society College is acknowledged for the research.

REFERENCES

1. P. F. Felzenszwalb, R. B. Girshick, D. Mcallester, and D. Ramanan, "Object detection with discriminatively trained part-based models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, p. 1627, 2010.
2. Object Detection with Deep Learning: A Review Zhong-Qiu Zhao, Member, IEEE, Peng Zheng, Shou-tao Xu, and Xindong Wu, Fellow, IEEE
3. LeCun, Y., Bengio, Y. and Hinton, G., 2015. Deep learning. nature, 521(7553), pp.436-444
4. D. G. Lowe, "Distinctive image features from scale-invariant keypoints," Int. J. of Comput. Vision,vol. 60, no. 2, pp. 91–110, 2004.
5. N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in CVPR, 2005.
6. R. Lienhart and J. Maydt, "An extended set of haar – like features for rapid object detection," in ICIP, 2002.
7. C. Cortes and V. Vapnik, "Support vector machine, " Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
8. Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," J. of Comput. & Sys. Sci., vol. 13, no. 5, pp. 663–671, 1997.
9. M. Pandey and S. Lazebnik, "Scene recognition and weakly supervised object localization with deformable part-based models," 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 1307-1314
10. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in NIPS, 2012.
11. R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in CVPR, 2014.
12. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in CVPR, 2016.
13. Mandal, M., Kumar, L.K. and Saran, M.S., 2020. MotionRec: A Unified Deep Framework for Moving Object Recognition. In The IEEE Winter Conference on Applications of Computer Vision (pp. 2734-2743).
14. R.Muralidharan, Dr.C.Chandrasekar, Object Recognition using SVM-KNN based on Geometric Moment Invariant,International Journal of Computer Trends and Technology-July to Aug Issue 2011 ISSN:2231-2803 http://www.internationaljournalssrg.org Page 215.
15. 3D Object Recognition Using Convolutional Neural Networks with Transfer Learning Between Input Channels Luís A. Alexandre
16. B. Yang, J. Yan, Z. Lei, and S. Z. Li, "Aggregate channel features for multi-view face detection," in IJCB, 2014.
17. X. Chen, K. Kundu, Y. Zhu, A. G. Berneshawi, H. Ma, S. Fidler, and R. Urtasun, "3d object proposals for accurate object class detection," in NIPS, 2015.
18. K. Kang, H. Li, T. Xiao, W. Ouyang, J. Yan, X. Liu, and X. Wang, "Object detection in videos with tubelet proposal networks," in CVPR, 2017.
19. S. Paisitkriangkrai, C. Shen, and A. van den Hengel, "Pedestrian detection with spatially pooled features and structured ensemble learning," IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, pp. 1243–1257, 2016.
20. W. Byeon, T. M. Breuel, F. Raue, and M. Liwicki, "Scene labeling with lstm recurrent neural networks," in CVPR, 2015

