International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

A REAL TIME YOLO HUMAN DETECTION IN FLOOD AFFECTED AREAS BASED ON VIDEO CONTENT ANALYSIS

Keerthana T[1], Kala L[2]

[1]PG Scholar, Department of Electronics and Communication Engineering, NSS College of Engineering, Palakkad, Kerala, India
[2]Associate Professor, Department of Electronics and Communication Engineering, NSS College of Engineering, Palakkad, Kerala, India

---

Abstract *Floods are becoming more frequent and severe natural disasters worldwide due to extreme climatic change. In addition to causing a huge economic damage (to the human property) they cause a substantial loss of human lives even leading to fatalities. Early detection is critical in providing a timely response to prevent damage to property and life .It is therefore crucial to use all available technologies, including Earth observation, in their prevention and mitigation. Person detection and tracking is a popular and still very active field of research in computer vision. There are many camera-based safety and security applications such as search and rescue, surveillance, driver assistance systems, or autonomous driving. Previous methods for flood detection uses specialized sensors or satellite imagery. In this paper, we propose a method for real time human detection based on video content analysis of feeds from surveillance cameras, which are more common and readily available nowadays. We demonstrate that YOLO is effective method and comparatively fast for recognition and localization in COCO Human dataset.*

Key Words: **Video *analytics; Computer vision; You Only Look Once (YOLO); COCO Human dataset; Convolutional Neural Network***

1. INTRODUCTION

In this new era, where technological advancements happen at a fast rate, we find solutions to a large number of problems faced by people around the world out of which natural calamities pose a major threat. Flooding is one of the major disasters occurring in various parts of the world. Because of global climate changes in recent years, rainfall has become comparatively heavy and rapid. It is the rise in the level of water leading to submerging of the land areas, especially the low lying areas, due to heavy downpour. It leads to loss of life and property. There has been government involvement in providing shelter to flood victims. Moreover, scientists have been researching for effective methods to deal with the damage caused. The rescue operations and basic amenities provided to victims must reach to heavily submerged area first. Here comes the importance of systems detecting the flood affected areas [1].These days, there are video surveillance systems everywhere.

The applications of convolutional neural networks (CNNs) have become fast-growing since [2] achieved significantly high accuracy. One of the major transferred tasks is object detection. The detection problem can be regarded as a task of labeling objects in an image with the correct class as well as predicting bounding boxes associated with real-valued confidence. Lots of researchers have proposed different structures of deep neural networks on this problem, such as Overfeat [3], DeepMultibox [4], Region CNN [5], YOLO [6], SSD [7], etc. Taking the advantage of rapidly developing methods, we here focus on a specific detection task: human detection in video streams Therefore, it is of high relevance to be able to identify and track the position of every person that is visible in a video stream recorded aligned with human vision. With the availability of large amounts of data, faster GPUs, and better and efficient algorithms, it is easily possible to train computers to detect and classify multiple objects within an image/video with high accuracy.

1.1 RELATED WORK

CNNs detection models work differently. For instance, RCNN (Girshick etal. 2014) extract potential bounding boxes using region proposal methods such as Selective Search (SS) and then classify these proposed bounding boxes with a CNN-based classifier. After classification, post-processing is used to refine the bounding boxes, eliminate duplicate detections, and rescore the boxes based on other objects in the scene. Despite the overall success of R-CNN, training an R-CNN model is expensive in terms of memory and time usage. By sharing computation of convolutional layers between region proposals for an image and replacing Selective Search (SS) with a neural network which is called Region Proposal Network(RPN), Fast R-CNN[9] (Girshick 2015) and Faster R-CNN [8](Ren et al. 2015) are able to achieve higher accuracies and better latencies overall. Instead of having a sequential pipeline of region proposals and object classification, YOLO[6] (Redmon et al. 2016a) method has formulated object detection as a single regression problem, straight from image pixels to bounding box coordinates and class probabilities. In this detection, a single convolutional network simultaneously predicts multiple bounding boxes and class probabilities for those boxes. YOLO trains on full images and directly optimizes detection performance. This leads a much lower latency

2. PROPOSED WORK

Firstly, the input video will be worked upon with a human detection algorithm, the detection algorithm will put a bounding box around the objects detected as people. The object detection algorithm, You Look Only Once (YOLO) [6] will be first used to detect the people in given input image/frame. In the past years, many face detection algorithms have been used, it could be either feature based or image based. Feature based algorithms detect face using facial features like edges, motion, point distribution models and many other. And the image based algorithm for face detection involve neural network, statistical approach. However YOLO out performs all of them. In terms of realtime performance, with the support of a GPU, YOLO-PC retrains a deep convolutional neural network to detect human at more than 40 fps (frames per second) [10]. Later, when the human is correctly detected in the input frame/image, it will be passed through a human tracking algorithm if necessary.
We used COCO human dataset for pretraining.

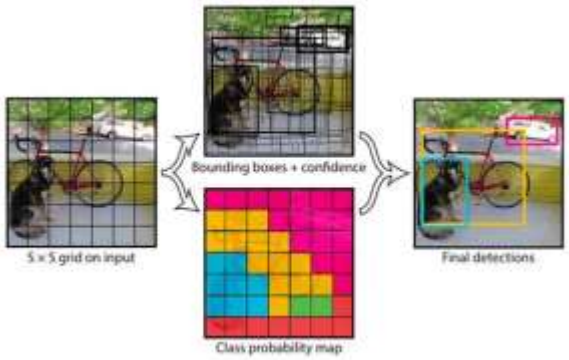

Fig -1: A simplified illustration of the YOLO object detector pipeline

2.1 Common Objects in Context (COCO)

Common Objects in Context (COCO) is developed by Microsoft and was introduced in 2015. The 2017 dataset contains over 120 000 images for training and validation, over 40 000 images for testing and 80 classes [11]. COCO is a large-scale object detection, segmentation, and captioning dataset.

2.2 Framework

Darknet[12] is a framework to train neural networks, it is open source and written in C/CUDA and serves as the basis for YOLO. Darknet is used as the framework for training YOLO, meaning it sets the architecture of the network. It is actually a complete neural network framework, so it really can be used for other objectives besides YOLO detection.

2.3 YOLO Algorithm

The YOLO [6] looks upon the input image/video as a single regression problem. Straight from image pixels to the bounding box is viewed as single problem. YOLO scans the image in one go. YOLO first takes an input image/video.

1. YOLO divides the image into S X S grids.
2. The bounding boxes are responsible for predicting up to 5 bounding boxes.
3. For each bounding box the class is predicted along with it prediction score.
4. The number of the bounding boxes we get will be 13 X 13 =169, and each grid producing up to 5 bounding boxes.

Therefore a total of 169 X 5 = 845 bounding boxes in total.

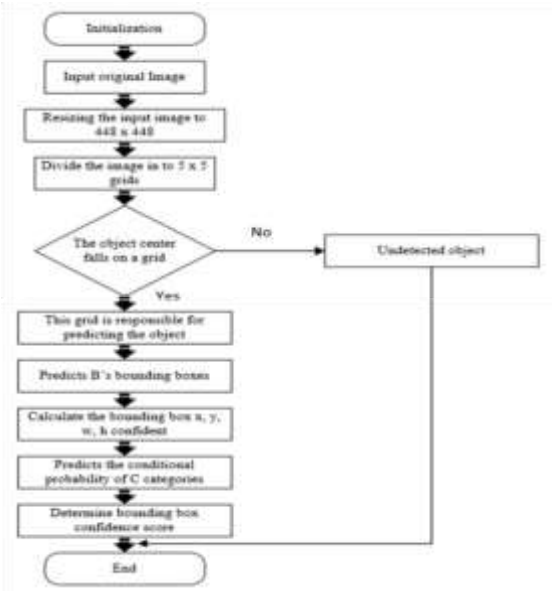

Fig -2: Flowchart of YOLO object detector

5. Most of the bounding boxes will have very low confidence score, and in the end only the bounding boxes crossing a certain threshold value would be included. For human detection threshold value of 30 sounds good enough.
6. Each bounding in Yolo predicts 5 parameters: x,y,w,h and confidence. The (x, y) coordinates represent the center of the box relative to the bounds of the grid cell. The width and height are predicted relative to the whole image. Finally the confidence prediction represents the IOU between the predicted box and any ground truth box
7. YOLO algorithm has a convolutional network used: convolutional 3X3 kernel and a max pooling of 2X2 kernal

4. EXPERIMENTAL RESULTS

OpenCV 4.0 version with Artificial Neural Networks (ANN) including CNN and YOLO Darknet-based Real-Time Object

Detector are used. The detection performance is evaluated on the annotated test frames from the COCO dataset. It is obvious that fine tuning the pre-trained network works better than the training the network from scratch. Detection examples are shown in Figures 3 and 4.

Fig -3: Multiple human detection with occlusion handling

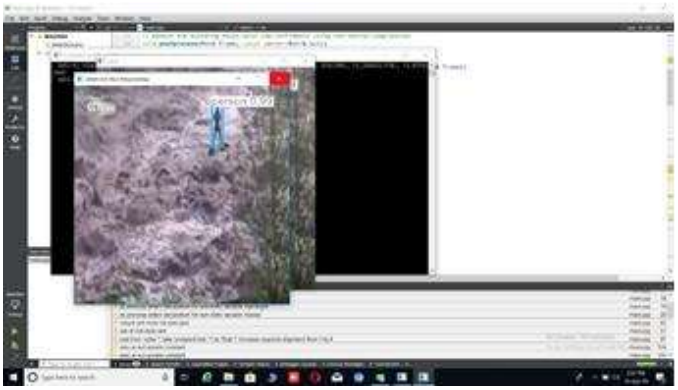

Fig -4: Single human detection with illumination changes

3. CONCLUSION

Object detection using deep learning and neural networks has taken some massive leaps in the past couple of years, and the field is very popular at the moment. Every month someone releases a new research paper, a new algorithm or a new solution for a certain problem. The main part of the goal was successfully implemented, a working application which utilizes neural network model for object detection. The detection and tracking performance is evaluated on the annotated test frames from the COCO dataset. A CNN based detection method called YOLO was also implemented to enhance the accuracy of human detection. YOLO showed a better performance in human detection

REFERENCES

1. Geetha, M., et al. "Detection and estimation of the extent of flood from crowd sourced images." International Conference on Communication and Signal Processing (ICCSP). IEEE, 2017.
2. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." Advances in neural information processing systems. 2012
3. Pierre Sermanet, David Eigen, Xiang Zhang, Michaël Mathieu, Rob Fergus, and Yann LeCun. Overfeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint ArXiv: 1312.6229, 2013
4. Erhan, Dumitru, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov. "Scalable object detection using deep neural networks." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2147-2154. 2014.
5. Girshick, Ross, et al. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2014.
6. Redmon, Joseph, et al. "You only look once: Unified,realtimeobjecdetection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
7. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016, October). Ssd: Single shot multibox detector. In *European conference on computer vision* (pp.21-37).Springer,October 2016.
8. Ren, Shaoqing, et al. "Faster r-cnn: Towards real-time object detection with region proposal networks." Advances in neural information processing systems. 2015.
9. Girshick, Ross. "Fast r-cnn." Proceedings of the IEEE international conference on computer vision 2015
10. Ren, Peiming, Wei Fang, and Soufiene Djahel. "A novel YOLO-Based real-time people counting approach." 2017 International Smart Cities Conference (ISC2). IEEE, 2017.
11. Ouaknine, Arthur. "Review of Deep Learning Algorithms for Object Detection." Medium. February 5 (2018): 2018.
12. https://pjreddie.com/darknet

BIOGRAPHIES

Keerthana T pursued Bachelor of Technology from Anna University in 2016. She is currently pursuing Master of Technology in Department of Electronics and Communication from APJ Abdul Kalam Technological University, since 2017. She has published review paper based on different methods for object detection and tracking in a reputed international journal. Her main research work focuses on deep

learning based human detection and tracking in different environments.

Kala L pursued Bachelor of Technology from Kerala University and Master of Technology from NIT Calicut. She is currently working as Associate Professor in Department of Electronics and Communication at NSS College of Engineering, Palakkad since 1991.
She has published many research papers in reputed international journals and conferences including IEEE. She has 27 years of teaching experience. Her interested areas include Digital Image Processing, Computer Vision, Deep Learning etc.