Person Detection in Surveillance videos

Freyal Shah, Zalak Shah, Khush Soni, Dhruvam Bhalodiya

Btech in Computer Science, Ahmedabad University

Email: [freyal.s, zalak.s, khush.s, bhalodiya.r]@ahduni.edu.in

Abstract— **Object detection is a subfield of computer vision that involves training models to recognize and locate specific objects in images or videos, such as people, animals, vehicles, and other objects of interest. Person detection is a specific case of object detection, where the object of interest is a human body. This task has numerous applications in various fields, including surveillance, robotics, and human-computer interaction. The main objective of person detection is to locate the presence of people accurately and efficiently in images or videos, regardless of variations in pose, appearance, and background. The aim of this report is to provide an overview of the different object detectors, challenges faced in detecting objects, implementation using YOLOv8 object detector and the results produced when tested against a test data.**

Keywords— Computer Vision, Object Detection, Deep Learning, Convolutional Neural Network, Training.

1. Introduction

Object detection deals with detecting objects of certain class such as humans, animals or cars in a digital image. Its goal is to develop models and techniques that will provide information about the location of the objects within the image. When the object detector is configured to detect only “person”, then this becomes a problem of person detection. Though the goal seems easy, the task of detecting objects is not so simple. There are numerous challenges that are faced while detecting objects. These include: occlusion, scaling, viewpoint and interclass variation. The detector that provides accurate results despite of the challenges, is the one that is used for object detection. There are a variety of object detectors that are employed for the purpose of object detection, most important of them include RCNN, faster RCNN, and YOLO.

1. LITERATURE SURVEY

Over the past few years, significant progress has been made, and numerous object detection algorithms have been proposed. In this literature review, some of the most prominent approaches to object detection are being explored.

**Viola-Jones Detector:** Viola-Jones detector is a classic object detection algorithm [1-2] that uses sliding window approach. This includes going through all the possible locations in the image and seeing if any window has an object. This algorithm is extensively used for face detection and has been extended to other object detection tasks. However, though the process was simple it was not a good idea to use this approach because the calculation was far beyond the computer's power.

**R-CNN (Regions with Convolutional Neural Network):** R-CNN is a popular object detection algorithm that uses region proposal methods like selective search [3] to identify potential object locations and then uses a convolutional neural network (CNN) to classify and refine the locations of the objects. R-CNN and its variants are still widely used in many applications. The major downside of this method is that there are a large number of overlapped proposals which leads to slow detection speed.

**Faster R-CNN**: Faster R-CNN is an extension of the R-CNN algorithm that uses a region proposal network to predict object locations, which makes it faster than R-CNN. It also uses a CNN to classify and refine the locations of the objects.

**YOLO (You Only Look Once):** YOLO is a real-time object detection algorithm [4] that works by dividing an image into a grid of cells and predicting the bounding boxes and class probabilities for each cell. It then uses a single neural network to predict object locations and their class probabilities. This algorithm is fast and accurate, making it a popular choice for real-time applications.

1. Dataset

The Crowd Human dataset has over 15,000 images with more than 350,000 labelled pedestrians. The images in the dataset were collected from a variety of sources, including surveillance cameras, internet images, and other public sources. The images depict a wide range of crowded scenes, such as streets, parks, and public events. Each image in the dataset is annotated with bounding boxes around each person in the scene, along with additional attributes such as occlusion, truncation, and crowd density.

In terms of density, on average there are around 22.6 persons per image in the dataset. Crowd Human contain people in unlimited poses in a wide range of domains to justify the factor of diversity.

1. IMPLEMENTATION

The first step in implementing an object detection model is the collection of dataset which contains annotated images having bounding boxes around the object of interest, here “person”. The dataset contains images with varying background, different orientations, and light conditions, which increases the robustness of the model that is being trained. Here, crowd dataset is selected for the detection [5]. After the dataset is collected, an appropriate detector is chosen to train the model by feeding the data to the pre-trained model chosen. Here, YOLOv8 is chosen to train the model pertaining to its ability to give state-of-art-performance. After the model is trained, it is tested using the test dataset to see its accuracy. The output images are the annotated images with bounding boxes around “person”. These images are manually compared with the actual images and the correctness of the model is checked.

1. RESULTS

The F1 confidence curve plots the F1 score as a function of the confidence threshold used by the model. The confidence threshold is the minimum level of confidence required for the model to classify a given instance as positive. As the confidence threshold increases, the precision of the model will increase while the recall will decrease. This means that the model will be more conservative in making positive classifications, but the classifications it does make will be more accurate. Conversely, as the confidence threshold decreases, the precision of the model will decrease while the recall will increase. This means that the model will be more liberal in making positive classifications, but some of the classifications it makes may be incorrect. By examining the F1 confidence curve, the optimal confidence threshold can be specified, here the confidence value that optimizes the precision and recall is 0.33 with the f1 score of 0.77.

1. CONCLUSIONS

Object detection is an crucial computer vision task that involves identifying objects of interest in an image or video stream and localizing them with bounding boxes. It is a challenging task, as it requires both accurate classification and precise localization of objects in the image. Moreover, object detection systems must be robust to variations in lighting, scale, orientation, and occlusion. Despite these challenges, object detection has made significant progress in recent years and is expected to continue to do so.

REFERENCES

1. P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” in CVPR, vol. 1. IEEE, 2001, pp. I–I.
2. P. Viola and M. J. Jones, “Robust real-time face detection,” International journal of computer vision, vol. 57, no. 2, pp. 137–154, 2004.
3. J. R. Uijlings, K. E. Van De Sande, T. Gevers, and A. W. Smeulders, “Selective search for object recognition,” International journal of computer vision, vol. 104, no. 2, pp. 154–171, 2013.
4. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You only look once: Unified, real-time object detection,” in CVPR, 2016, pp. 779–788.
5. Crowd Saliency Detection via Global Similarity Structure  
   M.K. Lim, V.J. Kok, C.C. Loy and C. S. Chan  
   Proceedings of 22nd International Conference on Pattern Recognition, pp. 3957-3962, 2014 ***(ICPR 2014)***