IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

OBJECT DETECTION IN SURVEILLANCE USING DEEP LEARNING

G. Aparanjini*, [1]P. HARI KIRAN,[2]B. VARSHITH KUMAR,[3]G. LITHISH KUMAR, [4]P. SAJEEVA RATHNA

(Assistant professor) [*] Department of CSE (AIML)

[1]Student, Department of CSE(AIML)

Nadimpalli Satya Narayana Raju Institute of Technology, Visakhapatnam, India

Abstract: Object detection in surveillance involves using computer vision algorithms to identify objects of interest in video footage. This technology has a wide range of applications, including security and safety monitoring, traffic management, and assistance for individuals with disabilities. The goal of object detection is to automatically locate and classify objects within a scene, such as cars, pedestrians, or specific individuals. This is typically achieved using machine learning algorithms that are trained on a large dataset of labeled images. The resulting model can then be applied to real-time video streams to detect and track objects of interest. Some challenges associated with object detection in surveillance include dealing with variations in lighting and camera angle, and ensuring the privacy and security of individuals captured by the surveillance system. Overall, object detection in surveillance has the potential to greatly enhance the capabilities of surveillance systems, making them more efficient and effective at identifying and responding to potential threats or situations of interest.

Index Terms - Deep learning technique, CNN (Convolutional neural network.), R-CNN (region based convolutional neural network)

INTRODUCTION

Object detection in surveillance involves using computer vision technology to automatically identify and locate objects of interest in video surveillance footage. This technology has become increasingly important in recent years, as the use of surveillance cameras has proliferated and the amount of footage generated has grown exponentially. Object detection algorithms are typically trained on large datasets of labeled images, which are used to teach the algorithm to recognize specific objects. The training process involves adjusting the parameters of the algorithm, such as the features it uses to represent objects and the criteria it uses to make decisions, in order to improve its performance. Once the algorithm has been trained, it can be applied to new footage to automatically detect and track objects in real time. Object detection in surveillance has a wide range of applications, including security and public safety, traffic monitoring, and retail analytics. In a security setting, object detection algorithms can be used to identify and track individuals, vehicles, or other objects of interest, and to generate alarms or alerts when unusual behavior is detected. For example, an object detection algorithm could be trained to recognize the presence of a person carrying a backpack, and to alert security personnel if the person enters a restricted area.

In a traffic monitoring setting, object detection algorithms can be used to track the movements of vehicles and pedestrians, and to generate statistics on traffic flow and congestion. This information can be used to optimize traffic signals and routes, and to improve the efficiency and safety of the transportation system. In a retail setting, object detection algorithms can be used to monitor customer behavior and to generate insights on shopping patterns and preferences. This information can be used to improve the layout of the store, to optimize the placement of products, and to target marketing efforts more effectively. Overall, object detection in surveillance is a powerful tool for enhancing situational awareness and enabling more effective and efficient surveillance operations. It is a rapidly growing field, and continues to evolve as new algorithms and technologies are developed.

NEED OF THE STUDY

The need for studying object detection in surveillance arises from the increasing use of surveillance cameras and the growing amount of footage generated by these cameras. Object detection algorithms are a key technology for automatically analyzing this footage and extracting valuable information. This information can be used to enhance security and public safety, to improve traffic management, and to gain insights into customer behavior, among other applications. Furthermore, the development of object detection algorithms is a rapidly growing field, and there is a need for ongoing research and development to improve the accuracy and efficiency of these algorithms. This research can lead to new breakthroughs and advances in the field, which can have wide-ranging impacts on the use of surveillance technology. In addition, the study of object detection in surveillance raises

important ethical and privacy concerns. As surveillance technology becomes more sophisticated and widespread, there is a need to carefully consider the implications of this technology for individual privacy and civil liberties. The study of object detection in surveillance can help to inform these discussions and to guide the development of ethical and responsible uses of this technology.

Data and Sources of Data

CIFAR-10 is one of the most widely used datasets in computer vision. The dataset is divided into 10 classes, each with 6000 low-resolution images, a total of 50'000 training images, and 10'000 test images. The data set CIFAR-10 is used primarily for research purposes.

The ImageNet dataset is one of the most popular image databases for computer vision applications. It provides over 14 million annotated images divided across 20’000 categories and is an open database that is free to researchers for non-commercial use.

MS COCO which stands for Common Objects in Context, is a large-scale image dataset published by Microsoft. It has an extensive collection of annotated image data specifically useful for image detection, segmentation, and captioning applications.

Theoretical framework

The theoretical framework for object detection in surveillance is rooted in the field of computer vision, which is concerned with enabling computers to interpret and understand visual data. This field involves a wide range of subfields, including image processing, pattern recognition, and machine learning, which are all relevant to the study of object detection in surveillance. At a high level, the theoretical framework for object detection in surveillance can be described as follows. First, the surveillance footage is preprocessed to extract relevant visual information, such as the location and appearance of objects of interest. This preprocessing step typically involves image segmentation, which divides the image into smaller regions that correspond to individual objects. Next, the preprocessed data is fed into an object detection algorithm, which is trained to recognize specific objects and to locate them within the image. The algorithm typically uses machine learning techniques to identify the presence of objects, and to generate bounding boxes that enclose the detected objects. Finally, the output of the object detection algorithm is post processed to generate the final result, which can take the form of a set of bounding boxes, object class labels, or tracking IDs. This output can then be used to trigger alarms, generate alerts, or provide input to other surveillance systems. Overall, the theoretical framework for object detection in surveillance involves a combination of image processing, pattern recognition, and machine learning techniques, which are applied to surveillance footage in order to automatically detect and track objects of interest.

RESEARCH METHODOLOGY

The research methodology used in object detection in surveillance typically involves a combination of theoretical analysis, experimental evaluation, and application-oriented research. Theoretical analysis typically involves the development of new algorithms and techniques for object detection, which are designed to improve the accuracy and efficiency of the detection process. This research often involves the use of mathematical models and computational techniques to analyze the performance of the algorithms and to identify potential improvements.

MODELS USED: -

Fast R-CNN

Fast R-CNN is a type of object detection algorithm that is designed to improve the speed and accuracy of the detection process. It works by first applying a convolution neural network (CNN) to the entire image to generate a set of region proposals, which are regions of the image that are likely to contain objects of interest. These proposals are then fed into a second CNN, which is trained to classify the objects in each proposal and to generate a set of bounding boxes that enclose the detected objects.
Finally, the bounding boxes are refined using a process called non-maximum suppression, which removes any overlaps or duplicates between the boxes. Overall, Fast R-CNN is a powerful and efficient object detection algorithm that has been widely adopted in a variety of applications.

Faster R-CNN

Faster R-CNN is a type of object detection algorithm that is designed to improve the speed and accuracy of the detection process. It is an extension of the Fast R-CNN algorithm, which uses a convolution neural network (CNN) to generate region proposals and to classify objects within those proposals. Faster R-CNN differs from Fast R-CNN in that it uses a single CNN to perform both the region proposal and classification stages. This is achieved using a technique called region proposal network (RPN), which is trained to predict the locations of objects in the image. The RPN is then combined with the classification CNN to generate the final set of bounding boxes that enclose the detected objects. Overall, Faster R-CNN is a powerful and efficient object detection algorithm that has been shown to outperform many other algorithms in terms of speed and accuracy. It has been widely adopted in a variety of applications, including security and public safety, traffic monitoring, and retail analytics.

Histogram of Oriented Gradients (HOG)

Histogram of Oriented Gradients (HOG) is a feature descriptor used in computer vision algorithms, including object detection. It is designed to capture the shape and orientation of objects in an image, by computing the distribution of gradient orientations within local regions of the image. The HOG descriptor is calculated by dividing the image into small cells, and then computing a histogram of gradient orientations within each cell. These histograms are then concatenated to form a single feature vector, which represents the distribution of gradient orientations within the entire image. The HOG descriptor can be used in a variety of object detection algorithms, including support vector machines (SVMs) and boosting-based methods. It has been shown to be effective at capturing the shape and orientation of objects, and has been widely used in a variety of applications, such as pedestrian detection and face recognition. Overall, HOG is a powerful and widely-used feature descriptor for object detection.

Region-based Convolution Neural Networks (R-CNN)

Region-based Convolution Neural Networks (R-CNN) is a type of object detection algorithm that uses a combination of region proposal and classification methods to detect and locate objects in images. It is a two-stage algorithm, in which the first stage generates a set of region proposals, which are regions of the image that are likely to contain objects of interest. These proposals are then fed into a second stage, which uses a convolution neural network (CNN) to classify the objects within each proposal and to generate a set of bounding boxes that enclose the detected objects. R-CNN was one of the first algorithms to achieve high accuracy on the challenging ImageNet object detection dataset, and it has been widely used in a variety of applications. However, it is computationally expensive, and has since been superseded by faster algorithms such as Fast R-CNN and Faster R-CNN.

Single Shot Detector (SSD)

Single Shot Detector (SSD) is a type of object detection algorithm that is designed to be fast and efficient. It is a single-stage algorithm, which means that it performs the detection and classification of objects in a single pass, without the need for a region proposal stage. SSD works by using a convolution neural network (CNN) to predict the locations and class probabilities of objects in the image. This prediction is performed at multiple scales and aspect ratios, in order to capture objects of different sizes and shapes. The output of the SSD algorithm is a set of bounding boxes that enclose the detected objects, along with associated class probabilities and other metadata. SSD has been shown to be effective at detecting objects in a wide range of applications, and has been widely adopted due to its speed and efficiency. It has also been extended to other tasks, such as facial landmark detection and instance segmentation. Overall, SSD is a powerful and widely-used object detection algorithm.

YOLO (You Only Look Once)

YOLO (You Only Look Once) is a type of object detection algorithm that is designed to be fast and accurate. It is a single-stage algorithm, which means that it performs the detection and classification of objects in a single pass, without the need for a region proposal stage. YOLO works by using a convolution neural network (CNN) to predict the locations and class probabilities of objects in the image. This prediction is performed on a grid of cells, with each cell responsible for predicting the presence and class of objects within its area. The output of the YOLO algorithm is a set of bounding boxes that enclose the detected objects, along with associated class probabilities and other metadata. YOLO has been shown to be effective at detecting objects in a wide range of applications, and has been widely adopted due to its speed and accuracy. It has also been extended to other tasks, such as instance segmentation and facial landmark detection. Overall, YOLO is a powerful and widely-used object detection algorithm.

CONCLUSION AND DISCUSSION

Object detection in surveillance is a powerful and widely-used technology that involves the use of computer vision algorithms to automatically identify and locate objects of interest in video surveillance footage. This technology has a wide range of applications, including security and public safety, traffic monitoring, and retail analytics. The development of object detection algorithms has been a rapidly growing field in recent years, and has led to significant advances in the accuracy and efficiency of the detection process. These advances have been driven by the use of machine learning techniques, such as convolution neural networks (CNNs), which have proven to be effective at capturing the shape and appearance of objects in images and video. Despite these advances, there remain many challenges and opportunities for further research and development in the field of object detection in surveillance. These include the development of more efficient algorithms, the integration of object detection with other surveillance technologies, and the careful consideration of the ethical and privacy implications of this technology. Overall, object detection in surveillance is a critical and rapidly evolving field, with the potential to have a significant impact on a wide range of applications. It is an area that continues to attract significant research and development efforts, and is likely to continue to evolve and advance in the coming years.

References

[1] Alexey Bochkovskiy, Chien-Yao Wang, and HongYuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. CoRR, abs/2004.10934, 2020

[2] Owei Cai and Nuno Vasconcelos. Cascade R-CNN: delving into high quality object detection. In 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018, pages 6154–6162. IEEE Computer Society, 2018

[3] Hao Chen and Abhinav Shrivastava. Group ensemble: Learning an ensemble of convnets in a single convnet. CoRR, abs/2007.00649, 2020.

[4] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet: Keypoint triplets for object detection. In 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019, pages 6568–6577. IEEE, 2019
[5] Jiyang Gao, Jiang Wang, Shengyang Dai, Li-Jia Li, and Ram Nevatia. NOTE-RCNN: noise tolerant ensemble RCNN for semi-supervised object detection. In 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019, pages 9507–9516. IEEE, 2019
[6] Yuqi Gong, Xuehui Yu, Yao Ding, Xiaoke Peng, Jian Zhao, and Zhenjun Han. Effective fusion factor in FPN for tiny object detection. In IEEE Winter Conference on Applications of Computer Vision, WACV 2021, Waikoloa, HI, USA, January 3-8, 2021, pages 1159–1167. IEEE, 2021
[7] Roman A. Solovyev, Weimin Wang, and Tatiana Gabruseva. Weighted boxes fusion: Ensembling boxes from different object detection models. Image Vis. Comput., 107:104117,2021.