OBJECT DETECTION USING CNN

Vinay L

prof .LIYA NAIJU

Student, Department of MCA, The Oxford College of Engineering (VTU), Bengaluru, India

Professor, Department of MCA, The Oxford College of Engineering (VTU), Bengaluru, India

Abstract

Detection can be difficult since there are all kinds of variations in orientation, lighting, background, and occlusion that can result in completely different images or videos of the very same object. Now with the advance of deep learning and neural network, we can finally tackle such problems without coming up with various heuristics real-time. The project “Object recognition using CNN for video streaming” detects objects efficiently based on CNN algorithm and apply the algorithm on image or video data. In this project, we develop a technique to identify an object considering the deep learning retrained model Mobile Net for Single Shot Multi-Box Detector (SSD). This algorithm is used for real-time detection, and for webcam feed to detect the purpose webcam which detects the object in a video stream. Therefore, we use an object detection module that can detect what is in the video stream. In order to implement the module, we combine the Mobile Net and the SSD framework for a fast and efficient deep learning-based method of object detection. The main purpose of our research is to elaborate the accuracy of an object detection method SSD and the importance of pre-trained deep learning model Mobile Net. The experimental results show that the Average Precision (AP) of the algorithm to detect different classes as car, person and chair is 99.76%, 97.76% and 71.07%, respectively. This improves the accuracy of behaviour detection at a processing speed which is required for the real- time detection and the requirements of daily monitoring indoor and outdoor.

Keyword: Sustainability, Environmental Impact, Detection Of Error, Database, Time Management ,Accurate.

1.INTRODUCTION

Blind to socioeconomic issues and inaccessible infrastructure. For a person who is blind, especially one who has lo stall eyesight, finding their way around is the hardest problem. Obviously, blind persons may move freely throughout their home without assistance because they are familiar with every room's layout .Finding nearby items might be challenging for blind individuals .Therefore, we created a REAL TIME object recognition system using CNN System. After reading a few publications in this field, we became interested in this topic. As a result, we are very driven to create a program that can identify objects in a real time context for video streaming.

Computer vision has some exciting challenges, like classifying images and identifying objects, both of which fall under the umbrella of object recognition. Convolutional neural networks, deep learning, and a rise in the parallelism processing power provided by graphics processing units (GPUs) have all made significant strides in recent years in the study of problems of this nature. The goal of the image classification problem is to select a label from a predetermined list of categories to apply to an input image. Despite being straightforward, this classification problem is crucial to computer vision because it has numerous practical applications, including labelling photos of skin cancer and using high-resolution images to find hidden objects.

Deep learning architectures with numerous specialized layers for automating the filtering and Use of feature extraction processes by today's most reliable object categorization and detection systems. A prediction is made, a correction is received, and the prediction mechanism is adjusted based on correction. At a high level, this is similar to how a human learns. Machine_learning_algorithms_like_linear_regression,support_vector_machines, and decision-trees all -have their own peculiarities in the learning process. The advantages of deep learning has introduced a fresh perspective on the issue, one that aimed to address past limitations by learning abstraction in data using a stratified description paradigm based on a nonlinear transformation. Deep learning's capacity to tackle semi supervised tasks is a crucial benefit.

CNN are widely used in deep learning-based algorithms, This enables understanding of these approaches' feature extraction stage. Convolution is a specific kind of linear operation that can be viewed in this sense as the straightforward application of a filter to a predetermined input. By adjusting the convolution's parameters, the same filter is applied to an input repeatedly to produce a feature map that shows the locations and intensities of any discovered features. As a result, the network can learn the ideal parameters to extract pertinent data from the database by adjusting itself to reduce error.

2. LITERATURE SURVEY

The description of object detection by Christian Szeged and others as a regression issue on object bounding box masks is straightforward but effective. It offers a multistate inference method that, when applied by a small number of network applications, yields high-resolution object detections at a reasonable price. Xiaog an gwan get offers a summary of deep learning with a focus on applications in object recognition, detection, and segmentation, which are key challenges for computer vision with numerous applications to still images and moving images. A unique object identification method is presented, and mean shift (MS) segmentation is used to better distinguish items. With the aid of depth data generated from stereo fixed number sliding window templates, vision is possible. As stated in Malay Sheet, Also possible is applying supervised learning to address the issue through method of the approach of imposing it with Decision trees or, SVM in deep learning. A thorough analysis of deep learning-based object identification frameworks is provided by ZhongQiuZhaoet, which addresses a number of sub-problems like confusion and low resolution brought on by differing degrees of RCNN modifications. Sandeep Kumara employs the easy net model, which enables prediction detection using just one network. The easy net version evaluates the entire image at testing time, allowing predictions to be informed by global context.

The thesis "A Survey on Object Detection and Tracking Algorithms" by Rupesh Kumar Rout focuses on various methods for object tracking, including multi-modal fusion, improving motion segmentation algorithms, and implementing the Block matching methodology. These techniques aim to enhance object detection and tracking by achieving better segmentation results and extracting more data using frame difference, Gaussian or mixture models, and background subtraction.

In another study by Dr. S. Padmavathi and Divya M., titled "Identification of the Movements of Humans in a Video," the Global GIST function is employed to detect human motions in videos. The study successfully tracks human movements using the Gist feature descriptor and skeleton detection. The authors express interest in future research to classify violence in conflict videos and differentiate between conflicts and non- conflicts involving multiple individuals.

3. Existing system

Object detection using Convolutional Neural Networks (CNN) has become a popular approach in computer vision tasks due to its ability to effectively learn features from images and identify objects within them. The existing system of object detection using CNN typically consists of the following key components:

1. Data Collection and Pre-processing: A large dataset of labelled images is collected to train the CNN model. These images are annotated with bounding boxes around the objects of interest. The data is then pre-processed to resize, normalize, and augment the images, which helps in improving the model's robustness and generalization.

2. Convolutional Neural Network Architecture: The CNN architecture is the core of the object detection system. Popular CNN architectures used for object detection include YOLO (You Only Look Once), SSD (Single Shot Multifood Detector), and Faster R-CNN. These architectures typically comprise a backbone network for feature extraction and additional layers for predicting object bounding boxes and class probabilities.

3. Feature Extraction: The backbone network of the CNN is responsible for extracting meaningful features from the input images. Commonly used backbone networks include VGG, Reset, and Mobile Net. These networks are pre-trained on large image classification datasets (e.g., Image Net) and then fine-tuned on the object detection dataset.

4. Bounding Box Prediction: After feature extraction, the CNN uses additional layers like convolutional layers and fully connected layers to predict bounding boxes around objects. These bounding boxes are represented as coordinates (x, y, width, height) and are combined with the class probabilities to define the detected objects.

5. Non-Maximum Suppression: Since multiple bounding boxes may be generated for a single object due to overlapping regions, non-maximum suppression (NMS) is applied to select the most accurate bounding box and eliminate redundant detections.

6. Training: The model is trained using the annotated dataset and optimization techniques like gradient descent with back propagation. The objective is to minimize the detection error by adjusting the model's parameters.

4. Proposed System

To address the challenges and improve the existing object detection system, a proposed system with several enhancements can be designed:

1. Efficient Backbone Network: To achieve real-time object detection, a lightweight and efficient backbone network like MobileNetV3 or Efficient Net can be employed. These models offer a good balance between speed and accuracy.

2. Anchor-Free Approaches: Moving away from anchor-based methods like Faster R-CNN, anchor-free approaches like Centre Net or FCOS can be considered. These methods predict the centre point of an object directly, simplifying the detection process.

3. Multi-Scale Feature Fusion: Introducing mechanisms to fuse features from different layers of the CNN can improve the detection of objects at various scales. Feature Pyramid Network (FPN) and PANet are examples of methods that facilitate multi-scale feature fusion.

4. Data Augmentation: Augmenting the training data with various transformations, such as rotation, flipping, and colour jittering, can help the model generalize better and improve its robustness.

5. Transfer Learning and Retraining: Leveraging transfer learning and retrained models can significantly reduce training time and enhance the performance of the object detection system.

6. Post-processing Techniques: Advanced post-processing techniques like Soft-NMS can be used instead of traditional NMS to improve the accuracy of object localization.

7. Hardware Acceleration: To achieve real-time performance on resource-constrained devices, hardware acceleration through techniques like model quantization or using specialized hardware (e.g., GPUs, TPUs) can be explored.

5. Functional requirements

Object detection using Convolutional Neural Networks (CNN) is a crucial computer vision task that involves identifying and localizing objects within images or videos. This technology has found applications in various

fields, including autonomous vehicles, surveillance, robotics, and augmented reality. In this context, functional requirements are essential to ensure the successful implementation and performance of an object detection system using CNN. Below are some key functional requirements:

1. Accurate Object Localization: The foremost requirement is to accurately detect and localize objects within the input images or frames. The CNN model should be able to draw bounding boxes around the objects with high precision, regardless of their size, orientation, or partial occlusion.

2. Support for Multiple Object Classes: The system must be able to detect multiple object classes simultaneously, such as cars, pedestrians, bicycles, traffic signs, and more. It should have the flexibility to detect a wide range of objects commonly encountered in real-world scenarios.

3. Real-Time Processing: For many applications like autonomous vehicles or robotics, real-time processing is critical. The object detection system must meet low-latency requirements to operate efficiently in dynamic environments.

4. Scalability: The system should be scalable to handle different image resolutions and video frame rates without sacrificing accuracy. This requirement ensures that the object detection model can adapt to various input sources.

5. Robustness to Variations: The object detection system should be robust to variations in lighting conditions, weather, and environmental factors to maintain reliable performance across diverse scenarios.

6. High Accuracy: The model should achieve a high level of accuracy in object detection to minimize false positives and false negatives, ensuring reliable decision-making in real-world applications.

7. Training and Re-Training Capabilities: It should allow periodic re-training to adapt the model to new or changing object classes, improving performance over time.

8. Support for Custom Object Classes: In some cases, the system may require the ability to add custom object classes relevant to specific applications, making it adaptable to specialized use cases.

9. Hardware Compatibility: The object detection system should be designed to run efficiently on various hardware platforms, including CPUs, GPUs, and specialized accelerators like TPUs (Tensor Processing Units).

6. CONCLUSION

This project is an efficient real-time deep learning base d framework to automate the process of monitoring and detecting. It uses frozen inference graph of the network, which has the calculated weights of different objects classified. GUI has also been provided for an effective and easy use with certain labels which are short descriptive and self-explanatory. The ability to recognise objects on either instantaneous detection or detecting objects from a video file. Each and every object that is detected is labelled with its appropriate name along with the accuracy of the detection around the bounded boxes is provided.

The project is a highly efficient and practical real-time deep learning-based framework that automates the process of monitoring and detecting objects. It utilizes a frozen inference graph, containing pre-calculated weights, to ensure fast and accurate object detection. The user friendly GUI enhances ease of use, presenting descriptive labels for detected objects and their corresponding accuracy scores. Whether for instantaneous detection from live video streams or processing object detection in pre-recorded video files, the framework excels, providing bounding boxes to precisely localize detected objects in the visual output.

7. FUTURE ENHANCEMENTS

This application is meant to be utilised in any working environment accuracy and precision are highly desired to serve the purpose. As mentioned, the proposed model is able to detect around 90 objects. As part of the future enhancements, the model will be custom trained and the other objects to increase its detection capability and by the support of transfer learning, the used network will be trained with other objects to increase the scope of objects the Mobile Net can detect.

8. REFERENCE

[1] E.-D. Company, C. Ghana, R.M. Hilo, and M. Gavrilescu, "Amount of plastic waste in the environment: ecological and health risks", Environmental Engineering and Management Journal, volume 3, issue 3, 2016, pages 675–687.

[2] The Food and Agriculture Organization, 2019. Annual Report. Accessible on the internet at the following URL: http://www.fao.org/platform-food-loss-waste/flw data/en/.

[3] G.T. Touché and T. Booth, "Evaluation of neglected bio-wastes potential with food-energy sanitation nexus", Journal of Cleaner Production, 242, 2020, 118547,.

https://doi.org/10.1016/j.jclepro.2019.118547
[4] A. Sorokowska An experimental study on attitudes towards food waste behaviour in children and adults was published in the Journal of Environmental Psychology volume 71, issue 2020, pages 101467. It was authored by M. Marcia, M. Misaim, M.M. Stefaczyk, and P.Sorokowski. The study was titled "Children older than five years do not approve of wasting food."