Computer Vision Application Analysis based on Object Detection

Chaitanya Chavan[1], Sanskruti Hembade[2], Gauri Jadhav[3], Parth Komalwad[4], Prerna Rawat[5]

Department of Computer Engineering

Genba Sopanrao Moze College of Engineering, Balewadi, Pune 45

--***---

ABSTRACT - Object detection is a computer vision method that allows for the identification and localization of specific classes of objects in images or videos. It goes beyond simple object classification and helps provide a better understanding of the object in question. Object detection has numerous applications, including automating business processes like inventory management in retail [1]. It can detect objects that occupy between 2% and 60% of an image's area and clusters of objects as a single entity. Additionally, it can localize objects at high speeds, typically greater than 15 frames per second. Vehicle detection is a crucial component in the development of autonomous vehicles, enabling them to identify and perceive objects in their environment. It involves identifying and locating vehicles in image or video frames and has various applications in surveillance and security systems. There are different techniques and models for object detection, including traditional image processing methods and modern deep learning networks. Traditional methods like Viola-Jones, SIFT, and histogram of oriented gradients do not require historical data for training and are unsupervised. Popular image processing tools like OpenCV can be used for these techniques. On the other hand, modern deep learning networks like CNN, RCNN, YOLO, ResNet, RetinaNet, and MANet are supervised and efficient for object detection.

Key Words: Deep learning, OpenCV, object detection

1. INTRODUCTION:

The development of image processing systems has shifted towards solving the problems of image processing itself rather than just developing the user interface. However, progress in solving tasks such as face recognition, vehicle detection, and remote and medical image analysis has been limited due to high costs and the need for trial and error. To automate the creation of software tools for solving these problems, a tool kit is required to support image analysis and recognition of previously unknown content, allowing ordinary programmers to develop effective applications. Object recognition includes tasks such as image classification, object localization, and object detection, where deep learning has emerged as a powerful tool for achieving state-of-the-art performance.

Vehicle detection is an important computer vision task used in various applications, but it is challenging due to variations in vehicle appearance, lighting conditions, and background clutter [2].

The project aims to provide an effective solution for vehicle detection in various applications, including traffic monitoring, intelligent transportation systems, and autonomous vehicles.

2. LITERATURE SURVEY:

SR. NO	Paper Name	Model	Summary
1	Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation[3]	Mask R-CNN ResNet-50 backbone architecture	Copy-Paste based data augmentation involves randomly scaling an image, cutting out a subset of the image, and pasting it elsewhere in the image. The authors demonstrate that this method improves the performance of existing instance segmentation models on the COCO and LVIS datasets.

2	Attention-based Joint Detection of Object and Semantic Part [4]	Mask R-CNN	The joint detection model for object and semantic parts combines an attention-based object detection module with a semantic segmentation module to achieve high accuracy in both tasks. The attention module is used to focus on relevant parts of the image, while the segmentation module is used to identify the semantic parts of the image.
3	Progressive End-to-End Object Detection in Crowded Scenes [5]	Sparse R-CNN	This paper presents a transformer-based progressive residual object detection model which is used to detect and segment objects in densely packed scenes. The results demonstrate that the proposed model achieves state-of-the-art performance on the Crowd-Pose dataset.
4	Leveraging Synthetic Data in Object Detection on Unmanned Aerial Vehicles [6]	RetinaNet feature pyramid network (FPN) backbone	The proposed method uses a combination of real and synthetic data to produce accurate object detection results, allowing UAVs to detect objects in challenging environments. The results demonstrate that the proposed method can achieve better performance than an object detector trained on real-world data alone.
5	PP-YOLOE: An evolved version of YOLO [7]	YOLO v4	PP-YOLOE is an evolved version of the YOLO (You Only Look Once) object detection model. The authors of the paper proposed a new architecture for the backbone network and introduced several improvements to the YOLOv4 architecture. They also introduced a new method called Progressive Parameter Pruning (PPP) to further improve the efficiency of the model. The resulting model achieves state-of-the-art performance on several object detection benchmarks while maintaining real-time processing speeds.
6	Waste Detection in Pomerania: Non-Profit Project For Detecting Waste In Environment [8]	YOLO V3	It describes the use of existing methods for waste detection, such as remote sensing, GIS, and image processing. Additionally, it discusses the development of a mobile application for waste detection and mapping. The project was successful in detecting various types of waste and was able to map the locations of waste with an accuracy of up to 80%.

7	PP-PicoDet: A Better Real-Time Object Detector on Mobile Devices [9]	PicoNet	The authors introduced an adaptive sampling strategy and a feature aggregation module called Pico-Fuse to improve the performance of PicoNet. The resulting object detector achieves state-of-the-art performance on several object detection benchmarks while maintaining real-time processing speeds on mobile devices.
8	Patch Refinement - Localized 3D Object Detection [10]	Patch Refinement	Patch Refinement works by first extracting patches from the image, then applying a convolutional neural network (CNN) to each patch to predict the object class and 3D bounding box. The predicted boxes are then refined by leveraging the contextual information from the surrounding patches. Experiments on the KITTI dataset show that the proposed method outperforms existing methods for localized 3D object detection.
9	A Normalized Gaussian Wasserstein Distance for Tiny Object Detection [11]	Normalized Gaussian Wasserstein Distance (NWD)	The NWD is a metric to measure the similarity between two images containing small objects. The NWD is shown to be more effective at detecting small objects than other existing methods, making it an attractive option for applications such as object detection in aerial imagery.
10	Object Detection with Spiking Neural Networks on Automotive Event Data[12]	Spiking Neural Network (LT-SNN-YOLOv2)	The proposed SNN model is trained on event data from automotive sensors and can detect objects in real-time. The authors validate their approach on multiple datasets, including image classification and object detection tasks

3. MODELS FOR VEHICLE DETECTION:

1) Faster R-CNN

Faster R-CNN is a deep learning model used for object detection that provides accurate predictions of object locations quickly. This model is designed as a single, end-to-end network that appears unified to the user. It outperforms its predecessor, Fast R-CNN, by incorporating a Region Proposal Network (RPN) with the CNN model. The RPN enables the generation of region proposals at virtually no cost by sharing the convolutional features of the entire image with the detection network.[13]

2) Sparse RCNN

Sparse RCNN is a method for detecting objects in images that performs well in terms of accuracy, speed and training convergence when compared to other established detectors on the COCO dataset[14]. It uses a one-to-one label assignment approach where the Hungarian algorithm is used to match each ground truth with only one positive sample.

3) Mask RCNN

Mask RCNN is a type of deep neural network designed to address instance segmentation tasks in computer vision[15]. It works by identifying the bounding box and class of objects in an image. Mask RCNN builds upon the Faster R-CNN architecture, which is a type of Region-Based Convolutional Neural Network.

4) YOLOv3

YOLOv3 is a cutting-edge algorithm that enables real-time detection of objects in videos, images, or live feeds. It is capable of identifying objects belonging to any of the 80 available categories, such as people, cars, and motorbikes, and producing bounding boxes for these objects from a single input image[16]. YOLOv3 stands out for its impressive accuracy of 28.2 mAP and speed, capable of achieving this accuracy in just 22 ms. In fact, YOLOv3 is as accurate as SSD but three times faster.

To improve upon its predecessors, YOLOv3 incorporates several enhancements, such as adding an objectness score to bounding box predictions, linking to the backbone network layers, and predicting at three different levels of granularity to boost performance on smaller objects. All these enhancements have contributed to making YOLOv3 a state-of-the-art object detection model that can be used in various applications, from surveillance systems to autonomous vehicles.

5) YOLOv4

YOLOv4 is an advanced model used for detecting objects in real-time and is the latest version of the YOLO family of models. It has been designed to cater to the specific requirements of vision engineers and developers working in custom domains[17], and its architecture is meant to ensure ease of use. YOLOv4 is a one-stage object detection model that outperforms its predecessor, YOLOv3, by incorporating various new techniques and modules from the research community. It offers the best possible speed and accuracy for object detection and can be trained on a single Graphics Processing Unit (GPU). The backbone of YOLOv4 is CSPDarknet53, complemented by a spatial pyramid pooling module, a PANet path-aggregation neck, and a YOLOv3 head. The code for YOLOv4 can be found on GitHub, and it is compatible with both Windows and Linux versions of Darknet.

6) RegNet

RegNet is a network design space made up of multiple stages, each containing multiple blocks. These stages form a stem (beginning), body (main part), and head (end). RegNet is not an architecture but rather a design space for networks. RegNet is a self-regulating network for image classification that can be easily integrated with any ResNet architecture[18]. It is a new type of regulated network that enhances the Squeeze-and-Excitation ResNet to demonstrate the method's generalization capabilities. RegNet is considered to be one of the most versatile network architectures for computer vision tasks.

7) RetinaNet

RetinaNet is a single-stage object detection model that employs a focal loss function to tackle class imbalance during training[19]. It integrates a Feature Pyramid Network (FPN) and adds classification and regression subnetworks to form an object detection model. RetinaNet is a robust model that uses a combination of Feature Pyramid Network and ResNet as its backbone.

8)   SSD (Single Shot Multibox Detector)

SSD (Single Shot Multibox Detector) is a cutting-edge object detection algorithm that can achieve accuracy levels like or even higher than Faster R-CNN while running much faster due to the absence of a region proposal network. It is a deep learning model that can detect objects in images or video sources and consists of two parts: the Backbone Model and the SSD Head. SSD uses feature maps from the entire input image to predict bounding boxes for each grid cell[20].

9)   Deformable DETR

Deformable DETR is an object detection technique that addresses the issues of slow convergence and high complexity found in DETR[21]. It achieves this by combining the sparse spatial sampling of deformable convolution with the relation modelling capability of Transformers. A deformable attention module is introduced that focuses on a small number of sampling locations to pre-filter key elements from all the feature map pixels.

4. APPLICATIONS:

Vehicle detection is used in a wide range of applications, such as traffic monitoring, road safety, automated driving, and Intelligent Transportation Systems (ITS). It is also used in surveillance, security, and law enforcement applications. In traffic monitoring, vehicle detection is used to monitor the flow of traffic and identify congested areas or areas where there is a high risk of accidents. In automated driving, vehicle detection is used to detect obstacles in the path of the vehicle and help the vehicle navigate safely. In surveillance and security, vehicle detection is used to detect suspicious vehicles and track their movements. Finally, in law enforcement applications, vehicle detection is used to identify stolen vehicles and detect violations of traffic laws.

5. CHALLENGES:

Vehicle detection is a challenging task due to the various changes in the background, the different sizes of vehicles, the presence of hard negative examples, and the low resolution of aerial images. Additionally, vehicle detection in night scenes is particularly challenging due to the lack of obvious vehicle features. To address these challenges, researchers have proposed various detection methods, such as R-CNNs, YOLOv3, and SSDs, as well as datasets such as XL-VIRAT and VisDrone.

6. CONCLUSIONS:

The papers discussed in this conversation delve into the subject of object detection models and how they can be improved for various applications. These include the detection of tiny objects, localizing 3D objects, and detecting objects in real-time using automotive sensors. The models used for this purpose are diverse and range from Mask R-CNN to YOLOv4, Sparse R-CNN, RetinaNet, PicoNet, Patch Refinement, and Spiking Neural Network (LT-SNN-YOLOv2). The techniques proposed for improving the models are equally fascinating, and include copy-paste based data augmentation, attention-based joint detection, transformer-based progressive residual object detection, synthetic data, and adaptive sampling with feature aggregation. In essence, these papers provide a wealth of information on the latest advancements in object detection and offer exciting solutions to the challenges faced by this field.

Future work in vehicle detection can focus on improving accuracy, robustness, and real-time processing of detection algorithms, as well as developing multi-object tracking algorithms, creating large-scale datasets, and using transfer learning techniques. These efforts can lead to more efficient, reliable, and adaptable vehicle detection systems with many potential applications in autonomous driving, traffic monitoring, and surveillance.

REFERENCES:

[1] https://en.wikipedia.org/wiki/Object_detection

[2] https://en.wikipedia.org/wiki/Vehicle_tracking_system

[3] "Simple Copy-Paste is a Strong Data Augmentation Method for Instance, Segmentation" Golnaz Ghiasi, Yin Cui, Aravind Srinivas, Rui Qian, Tsung-Yi Lin, Ekin D. Cubuk, Quoc V. Le, Barret Zoph arXiv:2012.07177v2 [cs.CV] 23 Jun 2021 CVPR 2021

[4] "Attention-based Joint Detection of Object and Semantic Part" Keval Morabia, Jatin Arora, Tara Vijaykumar arXiv:2007.02419v1 [cs.CV] 5 Jul 2020

[5] "Progressive End-to-End Object Detection in Crowded Scenes" Anlin Zheng, Yuang Zhang, Xiangyu Zhang, Xiaojuan Qi, Jian Sun CVPR 2022

[6] "Leveraging Synthetic Data in Object Detection on Unmanned Aerial Vehicles" Benjamin Kiefer, David Ott, Andreas Zell arXiv:2112.12252v1 [cs.CV] 22 Dec 2021

[7] "PP-YOLOE: An evolved version of YOLO" Shangliang Xu, Xinxin Wang, Wenyu Lv, Qinyao Chang, Cheng Cui, Kaipeng Deng, Guanzhong Wang, Qingqing Dang, Shengyu Wei, Yuning Du, Baohua Lai arXiv:2203.16250v3 [cs.CV] 12 Dec 2022

[8] "Waste Detection in Pomerania: Non-Profit Project for Detecting Waste in Environment" Sylwia Majchrowska, Agnieszka Mikołajczyk, Maria Ferlin, Zuzanna Klawikowska, Marta A. Plantykow, Arkadiusz Kwasigroch, Karol Majek arXiv:2105.06808v1 [cs.CV] 12 May 2021

[9] "PP-PicoDet: A Better Real-Time Object Detector on Mobile Devices" Guanghua Yu, Qinyao Chang, Wenyu Lv, Chang Xu, Cheng Cui, Wei Ji, Qingqing Dang, Kaipeng Deng, Guanzhong Wang, Yuning Du, Baohua Lai, Qiwen Liu, Xiaoguang Hu, dianhai yu, Yanjun Ma arXiv:2111.00902v1 [cs.CV] 1 Nov 2021

[10]   "Patch Refinement - Localized 3D Object Detection" Johannes Lehner, Andreas Mitterecker, Thomas Adler, Markus Hofmarcher, Bernhard Nessler, Sepp Hochreiter arXiv:1910.04093v1 [cs.CV] 9 Oct 2019

[11]   "A Normalized Gaussian Wasserstein Distance for Tiny Object Detection" Jinwang Wang, Chang Xu, Wen Yang, Lei Yu arXiv:2110.13389v2 [cs.CV] 14 Jun 2022

[12]   "Object Detection with Spiking Neural Networks on Automotive Event Data" Loïc Cordone, Benoît Miramond, Philippe Thierion arXiv:2205.04339v1 [cs.CV] 9 May 2022

[13]   "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun arXiv:1506.01497v3 [cs.CV] 6 Jan 2016, NeurIPS 2015

[14]   "Sparse R-CNN: End-to-End Object Detection with Learnable Proposals" Peize Sun, Rufeng Zhang, Yi Jiang, Tao Kong, Chenfeng Xu, Wei Zhan, Masayoshi Tomizuka, Lei Li, Zehuan Yuan, Changhu Wang, Ping Luo arXiv:2011.12450v2 [cs.CV] 26 Apr 2021

[15]   "Mask R-CNN" Kaiming He Georgia Gkioxari Piotr Dollar Ross Girshick arXiv:1703.06870v3 [cs.CV] 24 Jan 2018

[16]   "YOLOv3: An Incremental Improvement" Joseph Redmon Ali Farhadi arXiv:1804.02767v1 [cs.CV] 8 Apr 2018

[17]   "YOLOv4: Optimal Speed and Accuracy of Object Detection" Alexey Bochkovskiy Chien-Yao Wang Hong-Yuan Mark Liao arXiv:2004.10934v1 [cs.CV] 23 Apr 2020

[18]   "RegNet: Self-Regulated Network for Image Classification" Jing Xu, Yu Pan, Xinglin Pan, Steven Hoi, Fellow, IEEE, Zhang Yi, Fellow, IEEE, and Zenglin Xu arXiv:2101.00590v1 [eess. IV] 3 Jan 2021

[19]   "Focal Loss for Dense Object Detection" Tsung-Yi Lin Priya Goyal Ross Girshick Kaiming He Piotr Dollar arXiv:1708.02002v2 [cs.CV] 7 Feb 2018

[20]   "SSD: Single Shot MultiBox Detector" Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg

[21]   "Deformable Detr: Deformable Transformers for End-To-End Object Detection" Xizhou Zhu, Weijie Su, Lewei Lu, Bin Li, Xiaogang Wang, Jifeng Dai arXiv:2010.04159v4 [cs.CV] 18 Mar 2021