IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ENHANCING PEDESTRIAN SAFETY IN AUTONOMOUS VEHICLES THROUGH YOLOV5 DETECTION

[1]Athava Nanmathi J, [2] Dineshbabu K

[1]PG Scholar, [2]Associate Professor
Department of Electronics and Communication Engineering
Adhi College of Engineering and Technology, Anna University, Sankarapuram, India

Abstract **–** The main objective of the project is to detect pedestrians using the YOLOv5 deep learning model and enhance pedestrian safety. The goal of this innovative pedestrian detection and safety method is to address the shortcomings of the existing system, improving detection accuracy and enhancing pedestrian safety. Leveraging advancements from previous YOLO versions, the YOLOv5 model offers increased speed and accuracy while providing a unified framework for model training, enabling real-time object detection, which is gaining substantial attention. Detecting pedestrians is particularly critical for autonomous vehicles, yet it poses significant challenges due to variations in age, gender, clothing, lighting, backgrounds, and occlusion among pedestrians. The paper introduces a novel approach to pedestrian detection using a publicly available image dataset. Creating accurate annotations for this dataset is challenging, given the presence of images with extreme lighting conditions, such as direct sunlight, and unreliable intrinsic features for training. Therefore, we propose a customized YOLOv5 model trained with different loss and regularization methods to enhance its base accuracy. Based on our findings, we present a solution for addressing real-time pedestrian detection issues. These techniques hold promise for various applications, including Autonomous Vehicles and Advanced Driver Assistance Systems (ADAS).

Keywords **– Pedestrian Detection, YOLOv5, Deep Learning, SSD, CNN**

I. INTRODUCTION

Image processing is a transformative technique that converts visual information into digital data, enabling a range of operations aimed at enhancing images or extracting valuable insights from them. This technology revolves around the manipulation of images, such as video frames or photographs, to yield either improved visuals or specific characteristics relevant to the application. In essence, image processing treats images as two-dimensional signals and employs established signal processing techniques to analyze and manipulate them. In today's fast-evolving landscape, image processing finds extensive utility, spanning numerous sectors of industry and research. Its applications are diverse, addressing various aspects of businesses and advancing fields like engineering and computer science. Whether it's digitizing images through optical scanners or capturing moments through digital photography, image processing lays the foundation. The heart of image processing lies in tasks like data compression, image enhancement, and the detection of patterns that often elude the human eye, such as those found in satellite imagery. Ultimately, the culmination of these processes yields either an enhanced image or a comprehensive report, driven by the insights gleaned from rigorous image analysis. In this context, the intersection of image processing, deep learning, and pedestrian safety takes on particular significance, offering promising avenues for improving our understanding and response to the challenges faced by pedestrians in various environments.

II. LITERATURE REVIEW

A. New trends on moving object detection in video images captured by a moving camera

Comprehensive survey of the latest techniques for detecting moving objects in video sequences captured by mobile cameras. While extensive research exists for stationary cameras, this survey addresses the unique challenges of dynamic camera scenarios. The methods are classified into four categories: background subtraction, trajectory classification, low-rank and sparse matrix decomposition, and object tracking, with detailed explanations and notable improvements in each category. The paper also discusses challenges, concerns, performance metrics, and benchmark databases for evaluating these algorithms.

B. Scaled-YOLOv4: Scaling cross stage partial network

YOLOv4, an object detection neural network using the CSP approach, and demonstrates its versatility in scaling for both smaller and larger networks without compromising speed and accuracy. The proposed network scaling approach involves modifications in depth, width, resolution, and network structure. Notably, the YOLOv4-large model achieves outstanding results with a 55.5% average precision (73.4% AP50) on the MS COCO dataset, operating at approximately 16 FPS on Tesla V100. With test time augmentation, it achieves even higher accuracy at 56.0% AP (73.3% AP50), currently surpassing any published work on the COCO dataset. The YOLOv4-tiny model attains an impressive 22.0% AP (42.0% AP50) at a rapid speed of 443 FPS on RTX 2080Ti. Further optimization using TensorRT, batch size adjustments to 4, and FP16-precision pushes the YOLOv4-tiny to an astonishing 1774 FPS.

C. Deep learning strong parts for pedestrian detection

Recent advancements in pedestrian detection leverage Convolutional Neural Networks (ConvNets) to transfer learned features, effectively addressing pose, viewpoint, and lighting variations. However, these models often struggle when confronted with complex occlusions in pedestrian images. To combat this challenge, DeepParts is introduced as a novel approach, incorporating multiple part detectors. DeepParts offers notable advantages, including the ability to train effectively with weakly labeled data, accommodating low Intersection over Union (IoU) positive proposals, and empowering each part detector to function as a robust standalone detector, capable of accurate pedestrian detection even with partial information. This innovation significantly enhances pedestrian detection performance in scenarios characterized by occlusions.

III. PROPOSED SYSTEM ARCHITECTURE

A. Dataset Collection & Annotation:

Datasets are collected from public sources such as Pascal VOC and Kaggle. After collecting the dataset, annotation will be done using the makesense.ai online tool. Makesense.ai is a free-to-use online tool for labeling images for object detection.

B. Annotation into YOLOv5:

To train our custom model, we need to assemble a dataset of representative images with bounding box annotations around the objects that we want to detect. And then it will export into YOLOv5 format. For YOLOv5 format conversion, the online tool roboflow is used, which generates a correctly formatted custom dataset. Roboflow includes dataset splitting, image resizing, and image augmentation.

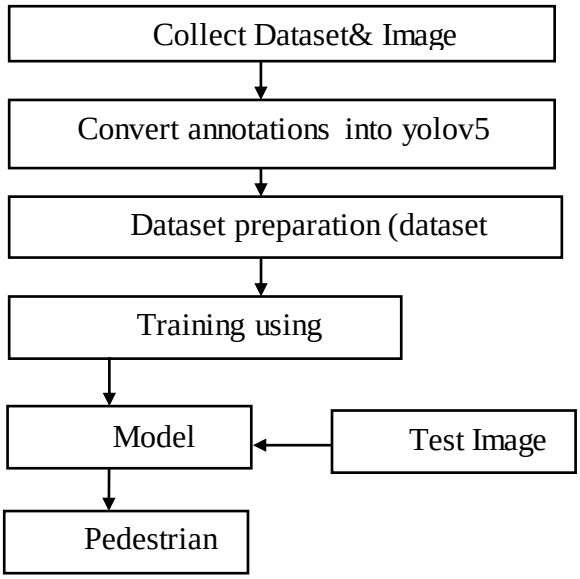

Fig. 3.1 Pedestrian Detection Architecture

C. Dataset Collection & Annotation:

Datasets are collected from public sources such as Pascal VOC and Kaggle. After collecting the dataset, annotation will be done using the makesense.ai online tool. Makesense.ai is a free-to-use online tool for labeling images for object detection.

D. Train YOLOv5:

We provide a unique deep learning framework for pedestrian detection in this work. In object detection and computer vision, deep learning models play a critical role. In this study, a YOLOv5 deep learning model is used to detect the pedestrian. In this network, the YOLOv5 model takes the extracted features as input from the CSPDarknet53 backbone model. This backbone is comprised of 53 layers. Here are the different building blocks of YOLOv5: Input, Backbone, Neck and Head

E. Pedestrian Detection:

Once the model has been trained, an input image will be provided and then given to the YOLOv5 detection module to detect the pedestrian based on the trained model.

IV. PROPOSED METHODOLOGY

Object detection in computer vision is crucial for identifying objects in images and videos. Common approaches like fast R-CNN, Retina-Net, and Single-Shot MultiBox Detector (SSD) have made significant progress in addressing data limitations and modeling challenges in this field. However, these methods typically require multiple steps to accurately detect objects. Although these approaches have addressed challenges related to data limitations and modeling in object detection, they often involve multiple algorithmic steps to achieve accurate results. This work introduces a YOLOv5 deep learning model for pedestrian detection. YOLO (You Only Look Once) is an algorithm that utilizes neural networks to enable real-time pedestrian detection. The proposed system encompasses the following processes: initially, gathering images from various sources and formatting them, followed by annotating them with ground-truth object bounding boxes. The collected and annotated dataset is subsequently converted into YOLOv5 format, involving dataset splitting and data augmentation. Next, we will develop a CNN-based model for pedestrian detection based on YOLOv5. In the training phase, images from the training set are used to train YOLOv5 for feature map extraction. Finally, pedestrian detection is applied to the given test image using the trained model.

V. RESEARCH AND DISCUSSION

The network architecture of YOLOv5, which includes three modules: (1) Backbone: CSPDarknet, (2) Neck: PANet, and (3) Head: YOLOv3 (anchor-based). The input data includes mosaic data augmentation and anchor box calculation are first input to CSPDarknet for feature extraction, and then fed to PANet to boost information flow. Finally, the Head consumes features from the Neck and outputs the results such as bounding boxes and class predictions. Hereby, SPP stands for Spatial Pyramid Pooling and Concat refers to concatenation. The network structure of YOLOv5 consists of three primary parts, as shown in Fig.2. Accordingly, the backbone is a CNN that aggregates and forms image features at different granularities. YOLOv5 adopts Cross-Stage Partial Networks (CSPNets) as its backbone to formulate image features. The network architecture of YOLOv5 consists of three modules: (1) Backbone: CSPDarknet, (2) Neck: PANet, and (3) Head: YOLOv3 (anchor-based). The input data includes mosaic data augmentation, and anchor box calculations are first input to CSPDarknet for feature extraction, and then fed to PANet to enhance information flow. Finally, the Head processes features from the Neck and outputs results such as bounding boxes and class predictions. "SPP" stands for Spatial Pyramid Pooling, and "Concat" refers to concatenation. The network structure of YOLOv5 comprises three primary components, as illustrated in Figure 5.1.

1. The backbone is a CNN that aggregates and forms image features at various granularities. YOLOv5 utilizes Cross-Stage Partial Networks (CSPNets) as its backbone to formulate image features. CSPNet addresses duplicate gradient problems in deeper CNNs, resulting in fewer model parameters and Floating-point Operations-per-Second (FLOPS), which improves inference speed, accuracy, and reduces the model size. The network also incorporates an SPP block after CSP to eliminate the fixed-size input image constraint. The SPP block computes feature maps from the entire image only once and then pools features from arbitrary regions (sub-images) to increase the receptive field and generate fixed-length representations for training the detectors.

2. The Neck in the model comprises a series of layers that combine image features and pass them forward for the detection stage. It employs a Path Aggregation Network (PANet) to enhance the information flowprocess. Specifically, the feature pyramid is enriched with accurate localization signals in lower layers through bottom-up path augmentation, shortening the information path between lower layers and the top feature.

3. The Head of the model is primarily responsible for the final detection stage. It utilizes feature anchor boxes and generates final output vectors containing class probabilities, objectness scores, and bounding box regression. YOLOv5 employs the same YOLOv3 (anchor-based) head for prediction. Additionally, YOLOv5 includes four models, each with varying memory storage sizes (parameters): YOLOv5s (the smallest, used in this paper), YOLOv5m (medium), YOLOvl (large), and YOLOv5x (extra large, the most prominent). All four models were trained on the MS COCO training dataset.

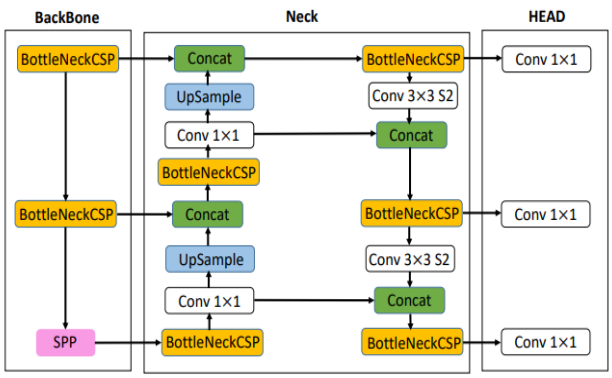

Fig. 5.1 The architecture of the YOLOv5 model

VI. ACKNOWLEDGEMENT

I would like to thank our guide and Head of the Department, Dr. Dineshbabu K for being ever supportive and helpful. He always pushed me to do better and showed immense faith in me.
In the end I would like to thank our parents for pushing us to achieve our best and motivating us to strive for excellence. I couldn't have gotten this far without their support.

VII. CONCLUSION

Pedestrians are among the paramount objects that autonomous vehicles must detect. Our YOLOv5-based pedestrian detection system stands as a transformative step in advancing pedestrian safety within the context of autonomous vehicles. By harnessing YOLOv5's cutting-edge features and harnessing richly diverse datasets, we've achieved a remarkable mean Average Precision (mAP) value of 95.2%. This level of precision is not just a technical achievement but a tangible enhancement of pedestrian safety. Swift and reliable pedestrian detection is an indispensable component in mitigating accidents and upholding the well-being of both pedestrians and vehicle occupants. Our research underscores the paramount significance of making pedestrian safety a top priority in autonomous vehicle technology. By doing so, we are actively contributing to safer streets, safeguarding lives, and propelling the evolution of autonomous driving technology towards a safer future.

REFERENCES

[1] Z. Q. Zhao, P. Zheng, S. T. Xu, and X. Wu, "Object Detection with Deep Learning: A Review," IEEE Transactions on Neural Networks and Learning Systems, vol. 30, no. 11. pp. 3212–3232, 2019, doi: 10.1109/TNNLS.2018.2876865.
[2] G. Chandan, A. Jain, and H. Jain, "Real Time Object Detection and Tracking Using Deep Learning and OpenCV" 2018 Int. Conf. Inven. Res. Comput. Appl., no. Icirca, pp. 1305–1308, 2018.
[3] M. Yazdi and T. Bouwmans, "New trends on moving object detection in video images captured by a moving
camera: A survey" Comput. Sci. Rev., vol. 28, pp. 157– 177, 2018
[4] Y. Wang, A. Bochkovskiy, and H.-Y.-M. Liao, "Scaled-YOLOv4: Scaling cross stage partial network" in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 13024–13033.
[5] Y. Tian, P. Luo, X. Wang, and X. Tang, "Deep learning strong parts for pedestrian detection," in Proceedings of the IEEE international conference on computer vision, 2015, pp. 1904–1912.
[6] C. Xie, P. Li, and Y. Sun, "Pedestrian Detection and Location Algorithm Based on Deep Learning," Proc. – 2019 Int. Conf. Intell. Transp. Big Data Smart City, ICITBS 2019, pp. 582–585, 2019, doi:10.1109/ICITBS.2019.00145.
[7] X. Jin, Z. Li, and H. Yang "Pedestrian Detection with YOLOv5 in Autonomous Driving Scenario" in 2021, IEEE,DOI:10.1109/CVCI54083.2021.9661188.
[8] R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers," in 2018 IEEE International Conference on Big Data (Big Data), 2018, pp. 2503–2510.
[9] Y. Jin, "Embedded Real-Time Pedestrian Detection System Using YOLO Optimized by LNN," no. June, pp. 1– 5, 2020.
10] G. Li, Y. Yang, and X. Qu, "Deep learning approaches on pedestrian detection in hazy weather," IEEE Trans. Ind. Electron., vol. 67, no. 10, pp. 8889–8899, 2020, doi: 10.1109/TIE.2019.2945295.
[11] Y. Jin, "Embedded Real-Time Pedestrian Detection System Using YOLO Optimized by LNN," no. June, pp. 1– 5, 2020.