Real-Time YOLO-based Heterogeneous Front Vehicles Detection

Masum Shah Junayed, Md Baharul Islam, Arezoo Sadeghzadeh, Tarkan Aydin

2021. 08. 27

INISTA-2021, Kocaeli, Turkey
Paper ID: 219

CONTENTS

BAU
BAHÇEŞEHİR ÜNİVERSİTESİ

IEEE
Celebrating 125 Years of Engineering the Future

Introduction

The Importance of Object Detection (OD)

Definition of Object Detection

Object detection in computer vision is referring to detecting the objects of a specific class (e.g. 'pedestrian', 'bicycle', 'car,' etc.) in images.

Applications of Video-based OD

Video-based object detection has been applied on many real-world applications such as vehicle detection, pedestrian detection, and traffic-sign detection.

Front Vehicle Detection

Why Vehicle Detection?

The perception of the complex road environment is a critical factor in autonomous driving, which has become the research focus in intelligent vehicles.

Why Vehicle Detection?

To having safe driving, intelligent vehicles should have a deep perception of the road environment and the vehicle behavior to determine the automatic driving path.

BAU BAHÇEŞEHİR ÜNİVERSİTESİ

IEEE Celebrating 125 Years of Engineering the Future

Introduction

Challenges of Video-based Vehicle Detection Approaches

Challenges of Vehicle Detection

1. **Limited to specific kinds of vehicles such as van, car, truck, and tram and not suitable for complex environments**

2. **Limited by the KITTI dataset, so their generalization ability needs to be improved.**

3. **Providing a real-time system is essential to take immediate action in autonomous driving for safety,**

4. **Being robust to the environment variations, e.g., illumination, background interference, and vehicle target changes (e.g., target occlusion)**

BAU BAHÇEŞEHİR ÜNİVERSİTESİ

IEEE Celebrating 125 Years of Engineering the Future

Introduction

The Main Goal of the Paper

A novel real-time model is proposed in this paper on a new dataset (DhakaAI dataset) containing a complex environment to overcome these challenges. This system is based on the YOLO model, which effectively detects and classifies various vehicles from both images and videos.

BAU BAHÇEŞEHİR ÜNİVERSİTESİ

IEEE Celebrating 125 Years of Engineering the Future

Literature Review

Object Detection

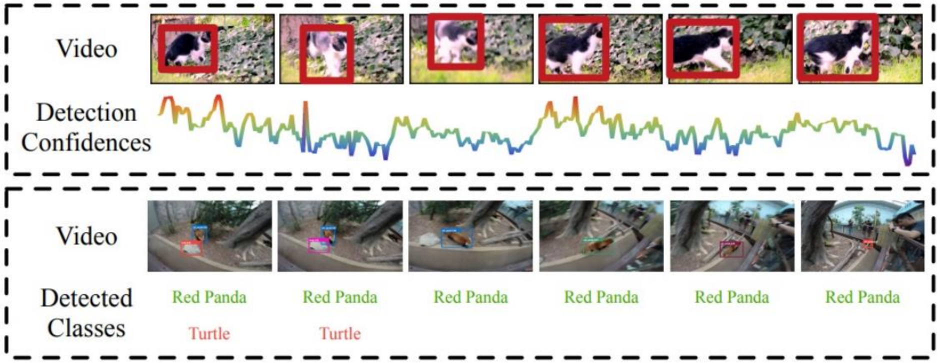
Static Image-based Approaches

1) Generating bounding boxes from the input image
2) Classifying each of these boxes as an object class
3) Applying post-processing techniques to improve the bounding boxes and their related class scores

Ineffective for videos as temporal and contextual information are involved

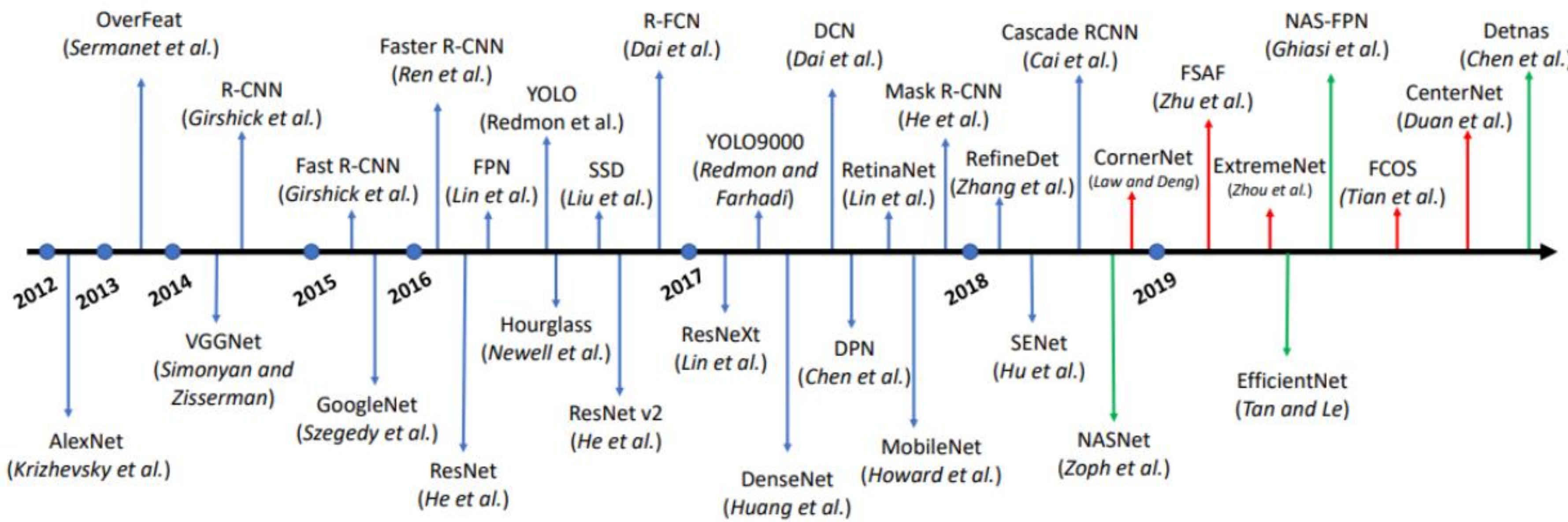
Video-based Approaches

- Introduced in 2015 by the ImageNet challenge
- Temporal information is considered in video object detection
- The bounding boxes of static object detection are considered sequences of bounding boxes



Literature Review

Milestone in object detection research based on deep CNNs since 2012



BAU BAHÇEŞEHİR ÜNİVERSİTESİ

IEEE Celebrating 125 Years of Engineering the Future

The Proposed Method

The Main Contributions

Contributions

01 **Employing DhakaAI dataset for the first time for vehicle detection to detect and classify 21 types of vehicles.**

02 **A modified YOLO-based feature extraction backbone is proposed to detect and classify the vehicles in front with high real-time accuracy (for both videos and images).**

03 **It is compared with four object detection approaches, namely Faster RCNN, SSD (single shot detector), RetinaNet, YOLOv3 are implemented on the DhakaAI dataset.**

BAU BAHÇEŞEHİR ÜNİVERSİTESİ

IEEE Celebrating 125 Years of Engineering the Future

The Proposed Method

DhakaAI Dataset

Dataset

1 A traffic dataset from front vehicles in the traffic crossroads of Dhaka

2 Images are obtained at different times of the day so that the light and shadow circumstances

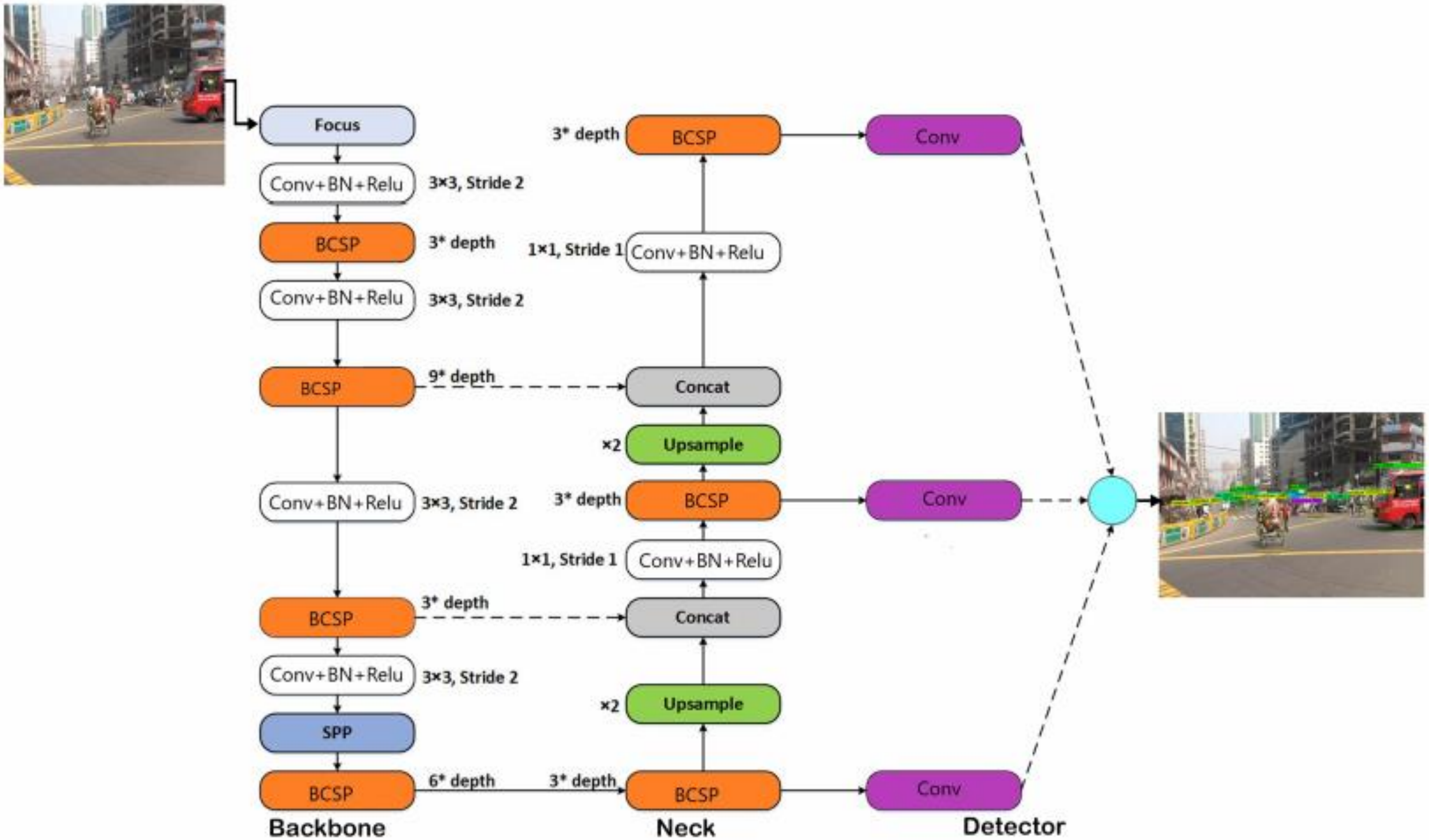
3 It contains various traffic flow conditions

4 The dataset has 21 different types of vehicles

5 The dataset has complex and crowded environment

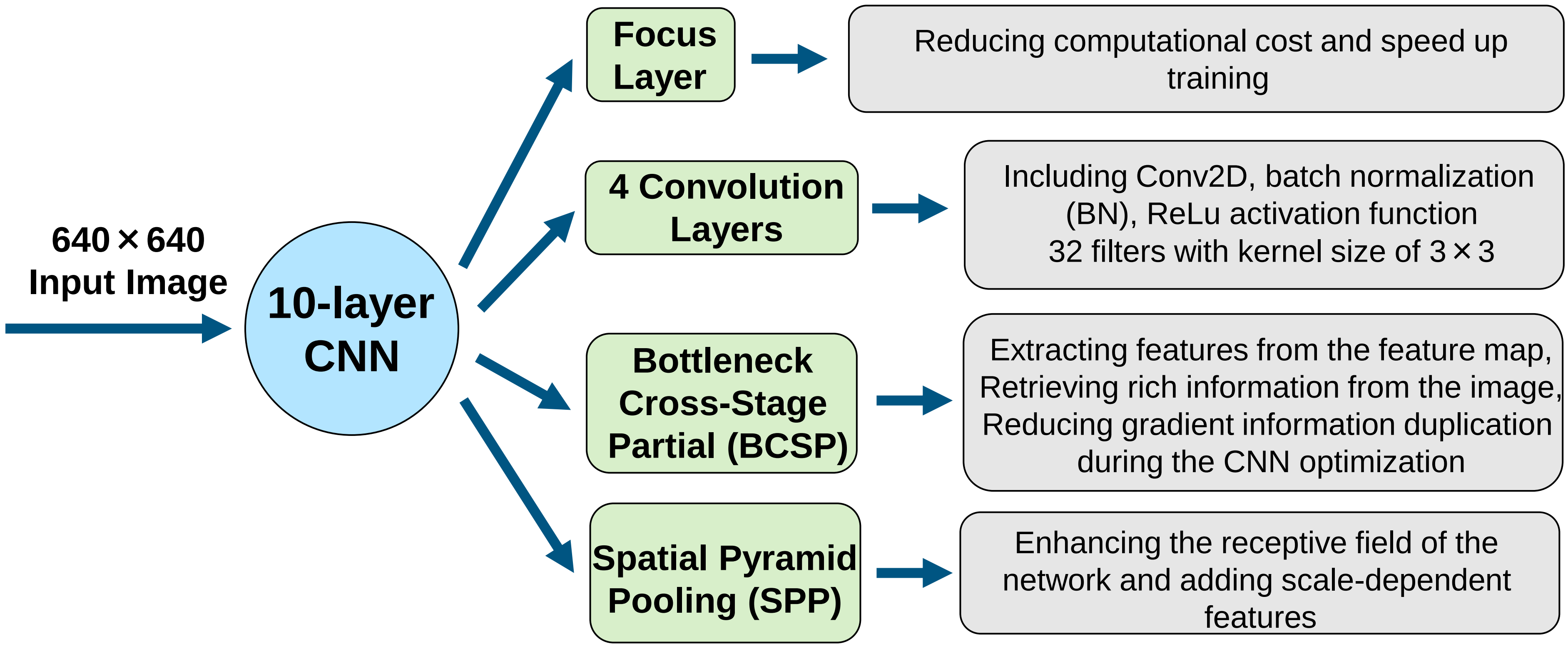
The Proposed Method

The Overall Flowchart of the Proposed Method



The Proposed Method

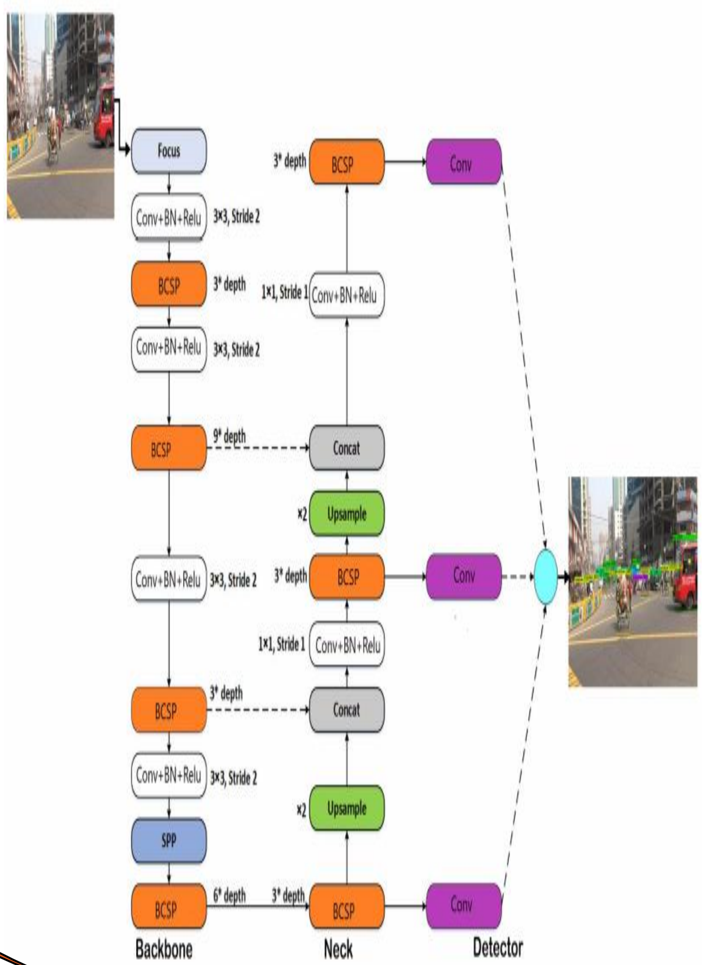
Backbone



The Proposed Method

Neck

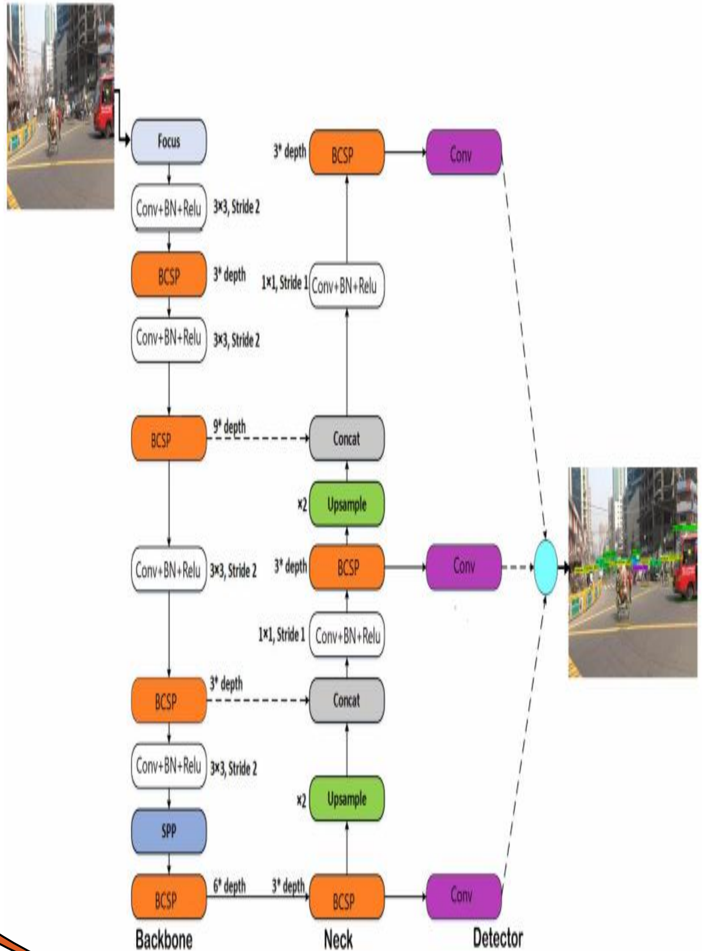
It is a sequential collection of aggregated image feature layers. The neck integrates and mixes image features to form feature pyramid networks (FPNs), mainly used to produce feature map characteristics.



The Proposed Method

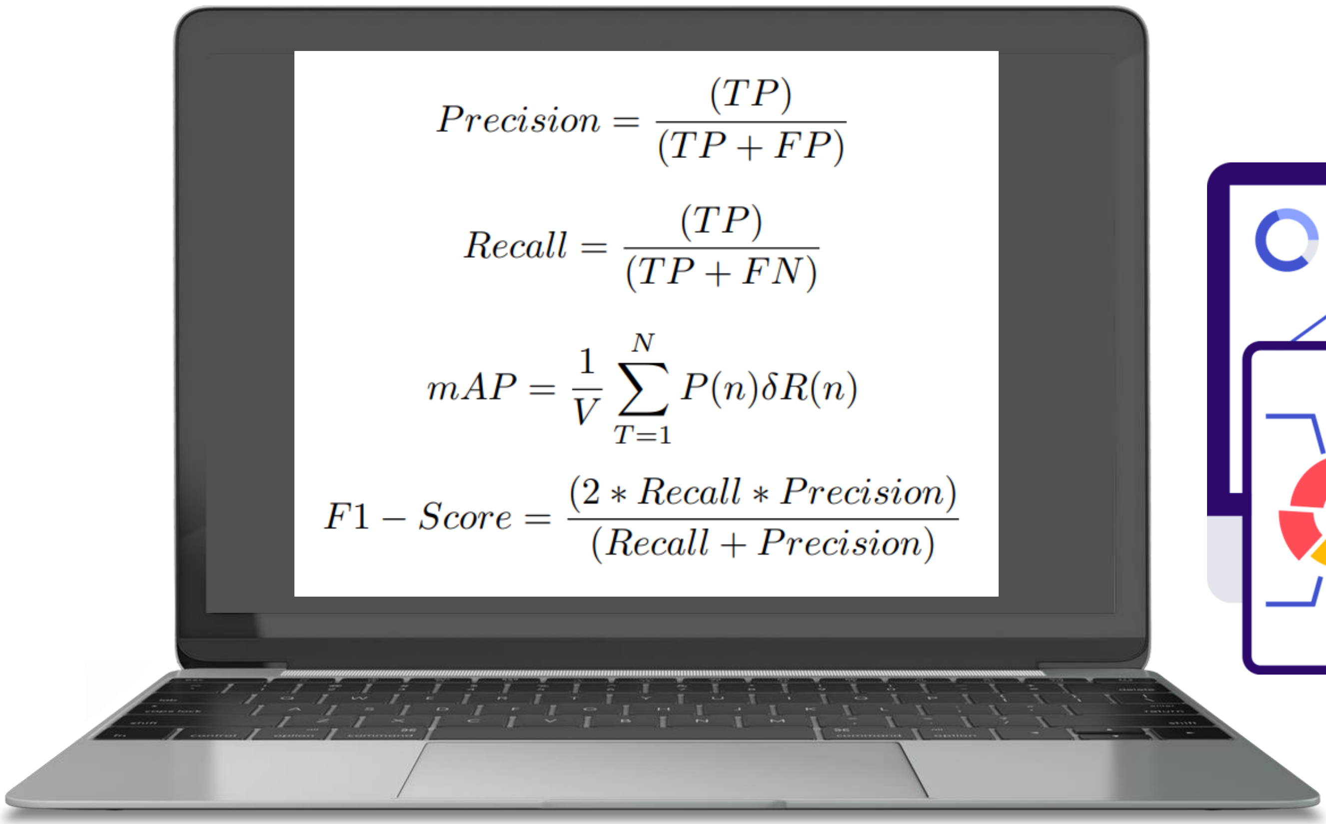
Detector

The final convolutional layers are used as detectors, which apply anchor boxes on the feature map generated from the previous layers. They produce a vector containing the category probability of the target vehicle, the object score, and the location of the bounding box around the vehicle.



Results and Comparison

Evaluation Metrics



$$Precision = \frac{(TP)}{(TP + FP)}$$

$$Recall = \frac{(TP)}{(TP + FN)}$$

$$mAP = \frac{1}{V}\sum_{T=1}^{N} P(n)\delta R(n)$$

$$F1 - Score = \frac{(2 * Recall * Precision)}{(Recall + Precision)}$$

Results and Comparison

Performance of the Proposed Method

	IOU=0.75			IOU=0.50		
Class	AP	TP	FP	AP	TP	FP
Ambulance	34.01	2	4	34.35	2	4
Army Vehicles	50.00	2	4	91.67	4	2
Auto Rickshaw	29.67	12	13	53.49	18	7
Bicycle	13.19	9	32	27.03	12	29
Bus	45.81	174	173	69.16	231	126
Car	46.14	265	318	72.32	358	225
Garbagevan	0.00	0	0	0.00	0	0
Human Haular	19.87	3	5	49.32	5	3
Minibus	31.62	3	4	31.62	3	4
Minvan	27.94	40	89	41.81	50	79
Motorbike	20.93	72	212	62.98	146	138
Pickup	34.55	41	54	54.86	53	42
Police car	0.00	0	0	0.00	0	0
Rickshaw	24.36	132	351	60.11	233	250
Scooter	0.00	0	0	0.00	0	0
Suv	26.73	32	164	45.89	42	154
Taxi	37.40	3	4	51.89	4	3
CNG	45.07	146	179	75.82	198	127
Truck	39.89	60	90	66.43	83	67
Van	25.54	16	47	43.92	23	40
WheelBarrow	13.01	2	13	26.25	4	11

Performance of the proposed system in terms of AP, TP, and FP for each type of the traffic vehicle classes based on two different IOU thresholds, i.e. 0.75 and 0.50.

IEEE
Celebrating 125 Years
of Engineering the Future

Results and Comparison

Performance Comparison

Method	$mAP_{0.75}$					
	Precision	Recall	F1-Score	Avg IOU	mAP	FPS
Faster RCNN [6]	31 .42	39.21	37.77	26.15	26.65	53.98
SSD [20]	33.19	41.56	36.93	28.67	23.89	51.51
RetinaNet [21]	34.54	43.16	38.67	29.08	25.63	54.17
YOLOv3 [22]	34.51	**43.22**	38.95	29.39	25.82	54.50
Proposed system	**35.99**	43.09	**40.11**	**30.82**	**26.94**	**55.83**

Method	$mAP_{0.50}$					
	Precision	Recall	F1-Score	Avg IOU	mAP	FPS
Faster RCNN [6]	50.02	59.81	55.43	40.80	40.91	54.89
SSD [20]	49.79	61.38	53.62	38.98	41.02	52.34
RetinaNet [21]	51.02	61.33	55.61	42.27	42.69	55.44
YOLOv3 [22]	51.68	61.37	55.46	**42.63**	43.87	55.98
Proposed system	**53.05**	**62.89**	**57.06**	41.37	**45.66**	**57.91**

Comparison between the proposed method and four common models of object detection in literature implemented on the same dataset of DhakaAI in terms of precision, recall, F1-score, Avg IOU, mAP, and FPS for two IOU threshold values of 0.5 and 0.75.

Results and Comparison

Graphical Representation of the Performance

Qualitative results of the proposed method on DhakaAI dataset at threshold of 0.5.

Conclusion

Characteristics and Future Works

Main Feature	**A recently published DhakaAI dataset with complex environments with different illuminations containing 21 types of vehicles and was employed to test the robustness and capability of the method.**
Main Feature	**A new YOLO-based model was proposed for vehicle detection which achieved an mAPof 45.66% and an FPS of 57.91 (at IOU threshold of 0.5).**
Future Work	**Providing further reductions in prediction time**
Future Work	**Making the method more robust to all kinds of vehicles by including additional vehicle types such as tractors, urban railcars, etc., in the database.**

BAU BAHÇEŞEHİR ÜNİVERSİTESİ

IEEE Celebrating 125 Years of Engineering the Future

Thank you
for your attention!