Journal of Computing & Biomedical Informatics
ISSN:2710-1606

Volume 09 Issue 01
2025

Research Article
https://doi.org/10.56979/901/2025

Leveraging Improved YOLOv10 for High-Performance License Plate Recognition in Challenging Mass Transit Environments

Faheem Mazhar[1*], Naeem Aslam[1], Muhammad Sajid[2], Ahmad Naeem[1], and Muhammad Fuzail[1]

[1]Department of Computer Science, NFC-IET, Multan, Pakistan.
[2]Department of Computer Science, Air University Islamabad, Multan Campus, Multan 60000, Pakistan.
[*]Corresponding Author: Faheem Mazhar. Email: faheemmazharbalouch@gmail.com

Received: April 20, 2025 **Accepted:** May 30, 2025

Abstract: Contemporary intelligent transportation systems significantly depend on automatic license plate identification for law enforcement, surveillance, and tollcollecting functions. Automatic license plate recognition is essential for vehicle management; yet, existing systems face challenges such as considerable angle deviations, subpar image quality, environmental disruptions, and barriers that render precise and rapid recognition in real-world contexts difficult in mass transit environments. Although current algorithms may identify license plates under optimal settings, their efficacy frequently diminishes in more intricate scenarios. This study presents a model for recognizing license plates that utilize enhanced YOLOv10 for plate detection and a distinct CNN for character recognition. The improved YOLOv10 algorithm greatly enhances the model's capacity to extract features and detect license plates efficiently. This research incorporates BiFPN, SEAM, and GCNet modules into the augmented YOLOv10 model. Additionally, we utilize extensive data augmentation to enhance the recognition of partially obscured plates in simulated adverse settings, augmenting the model's robustness in real-world applications. Experimental findings from our proprietary AOLP dataset indicate that our methodology attains a precision of 94.89%, a recall of 92.79%, an F1 score of 93.83%, and a mean average precision (mAP) of 94.10% for number plate detection, surpassing the baseline YOLOv10 by 3.22% in precision, 4.01% in recall, 3.83% in F1 score, and 0.56% in mAP. The R 2 value is 96.15%, accompanied by an RMSE of 4.01. The character recognition module of our suggested model attains an accuracy of 97.87% with a processing duration of 2.1 ms.

Keywords: Object Detection; YOLOv10; CNN; Object Recognition; License Plates

1. Introduction

Object detection [1], such as in practical applications, autonomous driving [2], robotic navigation [3], and object tracking [4], is an important aspect in the recognition of number plates. Recently, different authors utilized CNN for object detection in realtime applications [5–7]. Whereas these approaches demonstrate superior performance with standard images, they typically experience a reduction in detecting efficacy under unfavorable weather circumstances, particularly in fog, which is a prevalent occurrence in real-world situations [8, 9]. Researchers are encountering significant challenges due to the escalating traffic congestion. In the past, license plate identification has garnered significant interest from numerous academics [10]. Currently, the technologies for recognizing regular license plates are poorly developed, resulting in low resolution, pollution, inadequate lighting, and occlusion [10, 11]. To address these complex issues, many number plate recognition systems have been created, such as [5, 10–12]. However, these models still lack computational efficiency and precision. The detection network [13] identifies the position of the license plate in the image, while the recognition network interprets the sequence of characters on the license plate [14]. Recently, YOLO has been the leading paradigm in real-time object identification due to its efficient balance of processing expense and detection efficacy [1, 15, 16].

Domestic license plate identification technology has advanced quickly and steadily recently [17]. Deep learning models [18, 19] like YOLO and LPRNet can resolve the problem of reduced license plate identification under challenging conditions. The fact that foreign license plates can concurrently employ 0 and O and 1 and I as content makes recognition more challenging. Vehicle license plate recognition is made more difficult by these intricate vehicle license plate backgrounds, symbol combinations, and the variety of character combinations and abbreviations, which frequently necessitate high computational and time expenditures [20]. Researchers have investigated architectural designs, optimization aims, data augmentation methodologies, and other aspects of YOLOs, resulting in significant advancements [21, 22]. This work presents a comprehensive method integrating advanced object detection approaches for character segmentation to tackle the issues. The suggested work utilizes YOLOv10, an advanced object recognition technique [23], for exact character segmentation [24, 25]. This collaborative method seeks elevated precision and instantaneous processing skills crucial for effective ALPR applications [26].

There are far more vehicles on the road now, which encourages greater criminal activity [27]. Tracking cars is challenging due to their rapid growth, so the right authorities must do so. Automatic number plate recognition and vehicle make and model identification are becoming more and more necessary as the daily flow of cars increases [28, 29]. With the number of catastrophic incidents that happen every year, this application is also useful for tracking vehicle speeds. Thousands of surveillance cameras are positioned along the roadway, mostly for law enforcement and traffic control purposes [30]. It would be impossible to conduct continuous manual inspection, as doing so would involve enormous work and high expenses. Automatic visual interpretation enables the detection, tracking, and classification of all traffic. The make and registration information linked to the license plate are particularly significant concepts in this context. Another application involves tracking down specific cars after a crime, particularly when the license plate number is absent and only a description of the vehicle is available [31].

In these situations, it is necessary to visually identify the vehicle's make and model. By comparing the spotted automobile with its make and model, we can identify vehicles with stolen license plates [32]. This study introduces an enhanced YOLOv10 for license plate detection, utilizing a Cross Stage Partial Bottleneck-based CNN classifier for alphanumeric character recognition inside a multimodal framework. Enhanced YOLOv10 algorithm that greatly improves the model's feature extraction and detection skills. This research utilized BiFPN, SEAM, and GCNet. GCNet employs the global context block to record long range relationships in the image, improving the model's capacity to understand complex information with increased accuracy. This novel methodology integrates a comprehensive dataset augmentation technique to enhance the recognition of partial plates under simulated adverse settings, significantly enhancing the model's resilience in practical applications on a customized AOLP dataset. This study's primary novel contributions in addressing these challenges are as follows:

1. This research has tailored the AOLP dataset by incorporating numerous images from the subcontinent to enhance the diversity in the recognition of licensed number plates.
2. This research integrated advanced YOLOv10 with BiFPN, SEAM, and GCNet functionalities for object detection and segmentation.
3. Experiments conducted on a publicly accessible AOLP dataset revealed that the Enhanced YOLOv10 model attained enhanced detection accuracy.

1.1. Organization

The current research is structured as outlined: The section Background offers a comprehensive analysis of the background. The section Methodology defines the research technique employed in this study. The section Results and Discussion delineates the findings and interacts in discourse over these outcomes. The section Ablation Study offers a thorough summary of the insights and observations relevant to the proposed effort. The section Conclusion synthesizes the study's principal implications.

2. Literature Review

This section demonstrated the functionality of end-to-end object detection systems. Deep learning models offer an advanced approach to studying object detection and character recognition.

2.1. Real-Time Object Detection Systems

There are two primary ways for detecting license plates (LPD) in the literature: deep learning techniques [33] and classical techniques [34]. YOLO and SSD (Single Shot MultiBox Detector) are two popular real-time license plate detection designs [35]. Conversely, once the region containing the license plate has been recognized, the characters on the plate are identified using optical character recognition. We often utilize commercial software such as ABBYY FineReader and Tesseract OCR as tools [36]. Alternatively, it uses optical character recognition based on deep learning, including CNNs and RNNs, which have been utilized to improve the accuracy of character recognition [37]. The YOLO series [1, 38–41] is particularly prominent among mainstream options. YOLOv1, YOLOv2, and YOLOv3 delineate the conventional detection architecture comprising three components: backbone, neck, and head [38, 42]. YOLOv4 [43] and YOLOv5 [44] incorporate the CSPNet [45] architecture to supplant DarkNet [46], along with data augmentation techniques and improved PAN. [47] presents a C2F building block for effective feature extraction and fusion. [48] provides an improved gradient descent method to boost the effectiveness of multi-scale feature fusion. [49] incorporates GELAN to improve the architecture and PGI to boost the training process.

2.2. End to End Object Detection

The foundational structure of object detection is categorized as one-stage or twostage techniques, which have remained consistent for an extended period. [50, 51] formalized modern two-stage object detection by first introducing the RPN to extract regional features, followed by a subsequent stage to refine the predictions. Concurrently, [52, 53] introduced a one-stage object detector that directly regresses bounding boxes and predicts class probabilities from the convolutional features of the whole image within a single neural network, resulting in significant efficiency. Furthermore, [54] introduced a cascade methodology to develop a multi-stage detector. [55] subsequently improved performance by employing a sparse set of object suggestions instead of dense anchor priors. Alternatively, the Detection Transformer [56] was presented as an innovative method for object detection. Recently, Deformable DETR [57, 58] was introduced to mitigate these challenges by integrating deformable convolutions into the Transformer architecture, yielding significant improvements. This study presents an improved variant of YOLOv10 for the detection of license plates.

2.3. Deep learning Models

The authors utilized the TensorFlow framework in conjunction with the Keras deep learning package [27, 59]. A plate identification system utilizing a deep learning method is employed by the study [60]. This paper presents an innovative methodology for object detection utilizing improved YOLOv10 [1] and a CNN classifier [61] for character recognition. This study has customized the AOLP dataset by integrating various images from the subcontinent to improve the diversity in recognizing licensed number plates. Through a comprehensive background study, this approach demonstrates superior computational efficiency and processing power compared to existing models, as previously discussed. This paper highlights the transition from conventional methods to advanced machine learning and artificial intelligence methods. This demonstrates the continuous progress in ALPR via enhanced versions of YOLO and convolutional neural network classification capabilities.

3. Proposed Method

This study incorporated BiFPN, SEAM, and GCNet modules to improve the efficacy of license plate detection using the YOLOv10 model. The study seeks to improve recognition precision.

Table 1. Comparison with different Existing Techniques

Ref	Layer(L)	Parameter(M)	Key Feature	Use Case
ResNet[62]	50	25	Skip connections	Simple Tasks
EfficientNet[63]	237	66	Optimal Utilization	Remote devices
MobileNet[64]	28	3.2	Depthwise Convolution	Real-Time apps
NASNet[65]	1k	88	Cell-Based Framework	High Performance

Xception[66]	71	22	Depthwise convolutions	General purpose

3.1. Customized AOLP Dataset

The Application-Oriented License Plate dataset, sourced from [67], comprises 2,049 images captured under diverse locales, temporal contexts, traffic scenarios, and weather conditions. This study has customized the AOLP dataset by including a multitude of images from the subcontinent to augment the diversity in the recognition of licensed number plates. The entire dataset is divided into three subsets, each providing a substantial range of samples.

3.2. Preprocessing Methods

Preparing the data is important for analysis and predictive modeling because the quality and relevance of the data greatly affect the results and success of the models [68]. The preparation techniques utilized for the tailored AOLP dataset are detailed below:

• Resizing: The present work employed a resizing technique to adjust the dimensions of the input image, ensuring uniformity across all input images.

• Contrast Enhancement: This study employed a noise reduction technique to eliminate extraneous noise from the images.

• Color Space Conversion: This research applied a color space conversion approach to transform input photos to grayscale, therefore boosting processing efficiency and expediting the learning process, which in turn boosts performance.

3.3. Data Augmentation Methods

The original AOLP collection comprises 2,049 images. The dataset was insufficient, perhaps leading to imbalance and poor generalization. This technique yielded an improved compilation of 17,805 images.

• Random Horizontal Flip: This technique enables the model to learn from both the original image and its mirrored version, offering a broader array of training instances.

• Random Crop: Irrespective of the number plate's position inside the image, it acquires the ability to concentrate on the relevant region of interest (ROI) that encompasses it [68].

• Gaussian Blur: The application of a Gaussian blur reduces the model's sensitivity to extraneous noise, enabling it to concentrate on the fundamental structure and characteristics of the number plate images.

• Random Perspective: License plate images may display diverse shapes and abnormalities, leading to geometric distortions within the images. Utilizing a random perspective enhances the model's resilience to distortions.

Algorithm 1 Algorithm for feature extraction using Bi-Directional FPN

Input: I, License Plate Image.
Backbone Network: B YOLOv10, CNN.
Require: Number of BiFPN layers N.
Feature levels from Backbone $C = \{C_1, C_2, ..., C_L\}$.
Multi-scale feature maps $P = \{P_1, P_2, ..., P_L\}$.
Extract feature maps from Backbone: $\{F_1, F_2, ..., F_L\} \leftarrow B(I)$, where F_i corresponds to C_i.
Initialize BiFPN input levels: $P_i^0 \leftarrow F_i$ for $i = 1$ to L.
FOR $n = 1 : N$ **do**
  Forward Pathway:
  $P_1^{n,\mathrm{td}} \leftarrow P_1^{n-1}$
for $i = 2 : L$ **do**
  $P_i^{n,\mathrm{td}} \leftarrow \mathrm{Conv}(w_1 \cdot \mathrm{Resize}(P_{i-1}^{n,\mathrm{td}}) + w_2 \cdot P_i^{n-1})$
ENDFOR
  Backward Pathway:
  $P_L^n \leftarrow P_L^{n,\mathrm{td}}$
for $i = L - 1 : 1$ **do**
  $P_i^n \leftarrow \mathrm{Conv}(w_3 \cdot P_i^{n,\mathrm{td}} + w_4 \cdot \mathrm{Resize}(P_{i+1}^n))$
ENDFOR
ENDFOR
  Return $P = \{P_1^N, P_2^N, ..., P_L^N\}$

We consider an input image I, a backbone network B, the number of BiFPN layers N, and the feature levels C that originate from the backbone. The backbone network B analyzes the input image I to generate a collection of feature maps {F1, F2... FL} at various scales. Generally, it F1 exhibits the highest resolution (minimal downsampling), whereas it FL demonstrates the lowest resolution. The algorithm's foundation comprises N superimposed BiFPN layers. Every layer does a bi-directional feature amalgamation. Each level integrates the upsampled feature map from the preceding level with its corresponding original feature map. Trainable weights (w1, w2) are frequently employed for weighted feature integration. Beginning with the lowest resolution feature map, features are transmitted to higher resolutions. Each level integrates the downsampled feature map from the preceding level with the feature map from the top-down pathway at the corresponding level. Learnable weights (w3, w4) are prevalent in this context. The final output P is made up of several feature maps of different sizes taken from the last BiFPN layer, which can then be used for LPR tasks like detecting characters or recognizing sequences, as shown in Algorithm 1.

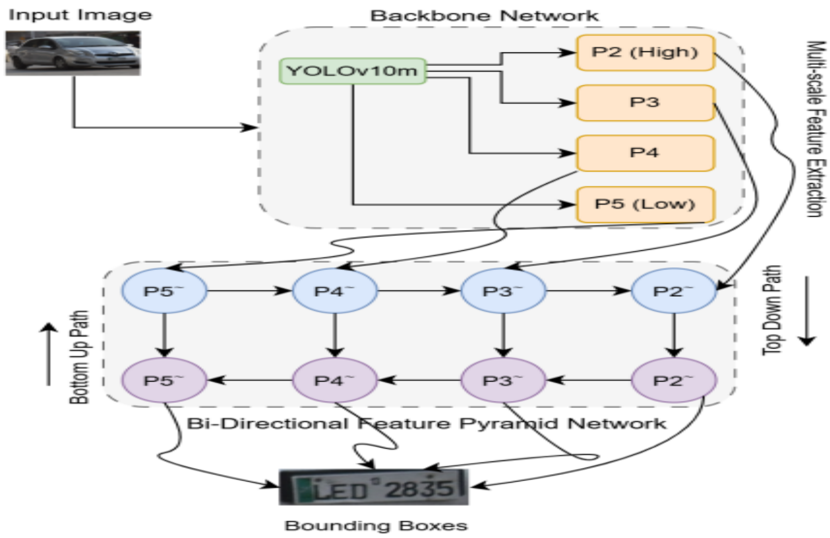

Figure 1. Implementation of Bi-directional Functionality Pyramid Network functionality in enhanced YOLOv10 architecture.

3.4. Bi-Directional FPN (Bi-DFPN)

BiFPN [69] facilitates efficient integration of multi-scale features via bidirectional pathways, enabling a more thorough capture of features across many layers, as shown in Figure 1. In conventional feature pyramid networks [70, 71], all input features are typically regarded as equivalent, resulting in the straightforward summation of features with disparate resolutions without accounting for their differing impacts on the output features. Algorithm 2 is essential to the Separated & Enhancement Attention Module

Algorithm 2 Algorithm for feature maps Separated & Enhancement Attention Module (SEAM)

Input: Feature Map $X \in \mathbb{R}^{C \times H \times W}$
Output: Enhanced Feature Map $Y \in \mathbb{R}^{C \times H \times W}$
Channel Attention Branch
$X_{gap} \leftarrow \text{GlobalAveragePooling}(X)$
$W_1 \in \mathbb{R}^{C \times C}, b_1 \in \mathbb{R}^{C}$
$W_2 \in \mathbb{R}^{C \times C}, b_2 \in \mathbb{R}^{C}$
$Z_c \leftarrow \sigma(\text{Linear}(X_{gap}, W_1, b_1) + \text{Linear}(X_{gap}, W_2, b_2))$
$S_c \leftarrow X \otimes Z_c$
Spatial Attention Branch
$X_{maxpool} \leftarrow \text{MaxPool}(X, \text{kernel\_size} = 3, \text{stride} = 1, \text{padding} = 1)$
$X_{avgpool} \leftarrow \text{AvgPool}(X, \text{kernel\_size} = 3, \text{stride} = 1, \text{padding} = 1)$
$X_{concat\_spatial} \leftarrow \text{Concat}([X_{maxpool}, X_{avgpool}], \text{dim} = 1)$
$W_s \in \mathbb{R}^{1 \times 1 \times 3 \times 2C}, b_s \in \mathbb{R}^{1 \times 1 \times H \times W}$
$M_s \leftarrow \sigma(\text{Conv}(X_{concat\_spatial}, W_s, b_s))$
$S_s \leftarrow X \otimes M_s$
$Y \leftarrow \text{Conv}(S_c + S_s, W_{final}, b_{final})$
Return Y

(SEAM), which features separate channel attention using two linear layers and then applies enhancement convolution after combining channel and spatial attention.

3.5. Separated & Enhancement Attention Module (SEAM)

The design mainly consists of depth-separable convolution, residual connections, and fully linked layers that integrate channel attention [72]. The arrangement is illustrated in Figure 2. This approach significantly reduces the number of FLOPs, lowering it from h×ω×k 2 timesC2 required by normal convolution to h×ω×k 2 timesC as shown in Equations 1, 2.

$$h \times \omega \times 2C \approx h \times \omega \times 2C + j^2 \times C \quad (1)$$

whereas conventional convolution required

$$h \times \omega \times 2C \approx h \times \omega \times 2C + j^2 \times C^2 \quad (2)$$

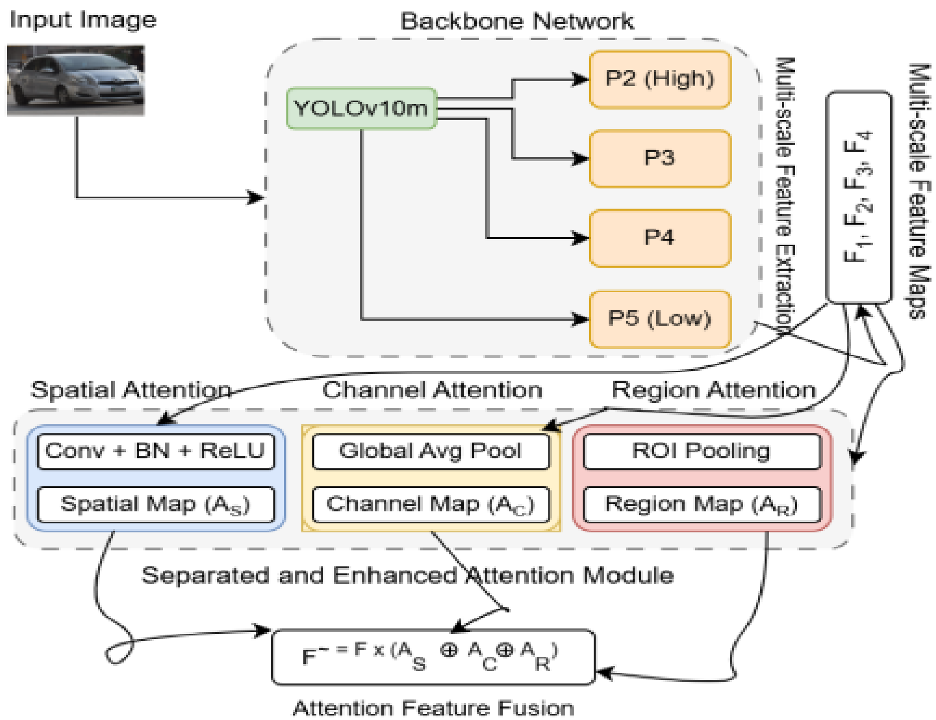

Figure 2. Implementation of the Separated and Enhanced Attention Module in the refined YOLOv10 Architecture.

This method decreases the parameter count while maintaining channel independence, hence enhancing computing efficiency and feature extraction accuracy, allowing SEAM to efficiently manage intricate features in complicated scenarios. The computational cost (FLOPs) of PConv is calculated by Equation 3.

$$h \times \omega \times j^2 \times C_l^2 \quad (3)$$

When rt = cj C = 1 4 , the FLOPs of PConv diminish to 1 16 of those of a standard convolution. PConv further necessitates less memory access as given by Equation 4,

$$h \times \omega \times 2C_j \approx h \times \omega \times 2C_j + l^2 \times C_j^2 \quad (4)$$

Algorithm 3 describes the Global Context Network (GCN) module, which aims to understand connections over long distances in the feature map's spatial layout. We utilized GCN because it enhances LPR by enabling the network to account for the complete context of the license plate while analyzing each local location.

3.6. Global Context Network (GCN)

Conventional CNNs predominantly depend on local receptive fields for convolutional processes during image processing. This approach is constrained by its ability to replicate only restricted regions of the image, leading to limited receptive fields, particularly when addressing objects with long-range correlations. GCNet substantially improves the model's feature representation capability by collecting long-range image relationships, as seen in Figure 3.

3.7. Enhancement Strategies Implemented on YOLOv10 Model

We improved feature extraction with the implementation of BiFPN, SEAM, and GCN. The recovered multi-scale features are fed into a Bi-Directional Feature Pyramid Network (BiFPN) to help combine features from different scales effectively. The output from BiFPN is further improved by one or more Separated & Enhancement Attention Modules (SEAM) to better focus on important details in both the channels and the space. We then transmit the attributes from SEAM across one or more Global Context Network (GCN) modules to capture extensive spatial dependencies. The processed multi-scale features, which may be concatenated, are subjected to global average pooling. A linear layer transforms the pooled features into logits that denote the probability of each character in the dictionary, together with a blank token, commonly utilized in sequence recognition tasks such as LPR with CTC loss; however, this pseudocode illustrates a more straightforward classification method for each output. We use SoftMax to calculate the probability.

Algorithm 3 Algorithm for capturing long-range dependencies using Global Context Network (GCN)

Input: Feature Map $X \in \mathbb{R}^{C \times H \times W}$
Output: Global Context Enhanced Feature Map $Y \in \mathbb{R}^{C \times H \times W}$
Global Context Modeling
$X_{flat} \leftarrow \text{Reshape}(X, (C, H \times W))$
$W_q \in \mathbb{R}^{C' \times C}$
$W_k \in \mathbb{R}^{C' \times C}$
$W_v \in \mathbb{R}^{C \times C}$
$Q \leftarrow W_q X_{flat}$
$K \leftarrow W_k X_{flat}$
$V \leftarrow W_v X_{flat}$
$S \leftarrow \text{Softmax}((Q^T K)/\sqrt{C'})$
$Context \leftarrow VS$
$Context_{reshape} \leftarrow \text{Reshape}(Context, (C, H, W))$
Feature Fusion
$W_c \in \mathbb{R}^{C \times C}, b_c \in \mathbb{R}^C$
$Y \leftarrow \text{ReLU}(\text{BatchNorm}(\text{Conv}(Concat([X, Context_{reshape}], \text{dim} = 1))))$
$Y \leftarrow \text{ReLU}(\text{BatchNorm}(W_c \cdot X + Context_{reshape} + b_c))$
Return Y

3.8. Focal Cross-Entropy Loss (FCE Loss)

The focal loss, through the incorporation of a focusing parameter γ, modulates the impact of each sample on the loss according to classification difficulty, diminishing the effect of easily classified cases while enhancing that of more challenging ones. Focal loss can be calculated by Equation 5.

$$FT_{pt} = -\alpha \times [1 - pt]^{\gamma} \times log_{pt} \quad (5)$$

Where pt estimates the probability of each class α is a weighted coefficient for each category, γ is a parameter for adjusting the rate of down-weighting easy instances.

3.9. Performance Evaluation Metrics

In the subsequent experiment, this study assesses the precision of number plate identification utilizing the following metrics:

• Precision: It is a crucial statistic for evaluating the efficacy of classification and object identification models. The formula is delineated by Equation 6:

$$P_r = \frac{TP}{TP + FP} \quad (6)$$

• Recall: This measure evaluates classification and object detection models by examining their ability to accurately recognize all pertinent instances within a dataset. It can be calculated by Equation 7.

$$R_e = \frac{TP}{TP + FN} \quad (7)$$

• F1 Score: It signifies the harmonic mean of precision and recall. It can be calculated by Equation 8.

$$F_1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

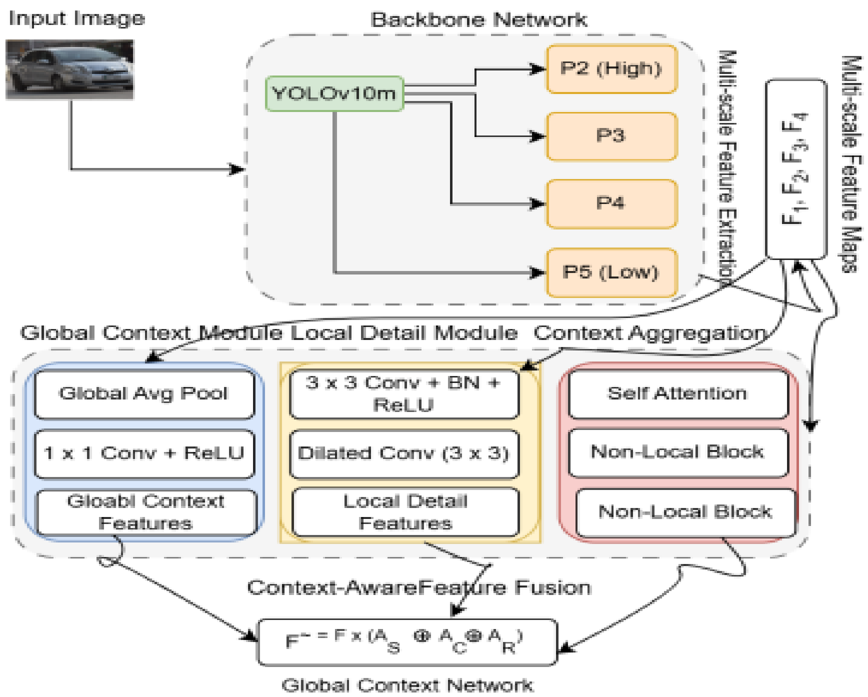

Figure 3. Feature extraction process of the Global Context Network functionality in the enhanced YOLOv10 architecture.

• R Square: R2 quantifies the extent of regression, demonstrating the correlation between anticipated number plate characters. It is calculated by Equation 9.

$$R^2 = \frac{\sum_{k=1}^{P_t}(t_k - P_k)^2}{\sum_{k=1}^{P_t}(t_k - \bar{t_k})^2} \quad (9)$$

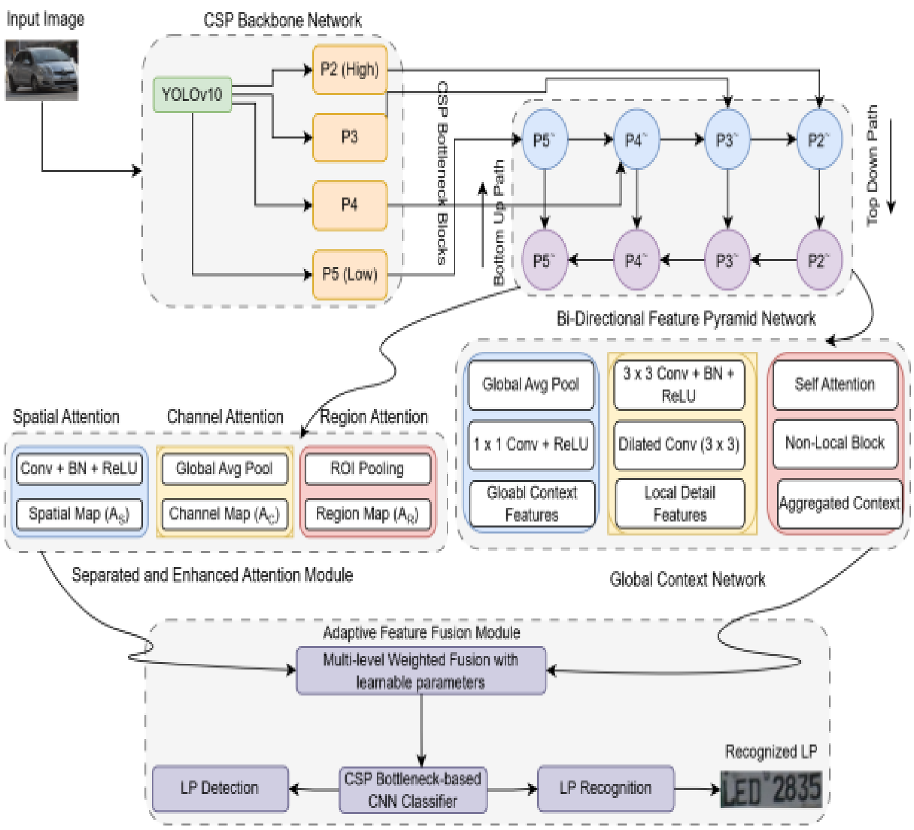

Figure 4. Character recognition via a CSPBottleneck-based CNN classifier, employing a convolutional neural network architecture.

4. Results and Discussion

This section offers a thorough description of the experimental methodology and training protocols. The best parameter configurations are as given in Table 2.

Table 2. Model Training Hyperparameters

Parameter	Value
Learning-Rate	1x10-3
Input image resolution	640x640
Weight-decay	0.005
Optimizer	SGD
No of Epoch	100
Batch Size	32
No of Iteration	1500

4.1. Model Training

To mitigate the issue of overfitting, the current research implemented various solutions. The current research partitioned the original and enriched dataset of 17,805 images using a standard ratio of 80:10:10 for the model. A total of 17,805 images were utilized for training, with 1,780 images allocated for the validation set and 3,561 images designated for the testing set. The YOLOv10 model served as a pre-training framework, facilitating the acquisition of generic features to improve the model's initial performance. Furthermore, throughout the training phase, this study employed the early stopping strategy, which halts training when the validation set performance fails to increase over ten consecutive epochs, to mitigate the risk of overfitting on the training set. The comparison is seen in Figure 5, and quantitative results are given in Table 3. BiFPN markedly enhanced detection accuracy by effective multi-scale and weighted feature fusion techniques. This dual attention method enhances the model's accuracy in detecting number plates in intricate traffic conditions.

4.2. CNN Classifier for Character Identification

The character segmentation algorithm examines cropped license plate images, whereas the character recognition technique utilizes these segmented images. Figure 6 illustrates the segmentation outcomes of YOLOv10 and the optimized YOLOv10 method.

Table 3. Comparative Evaluation of Proposed Model

Technique	Precision	Recall	F1 Score	mAP50	R2	RMSE
SSD	89.20	55.46	68.0	71.75	17.0	19.87
Faster R-CNN	68.57	69.0	71.75	69.64	56.0	14.52
YOLOv8	91.22	87.26	89.0	92.58	88.0	7.58
YOLOv9		55.46	89.0	93.47	89.0	7.26
YOLOv10	91.67	88.78	90.0	93.54	94.0	5.45
Proposed	94.89	92.79	93.12	94.10	96.15	4.01

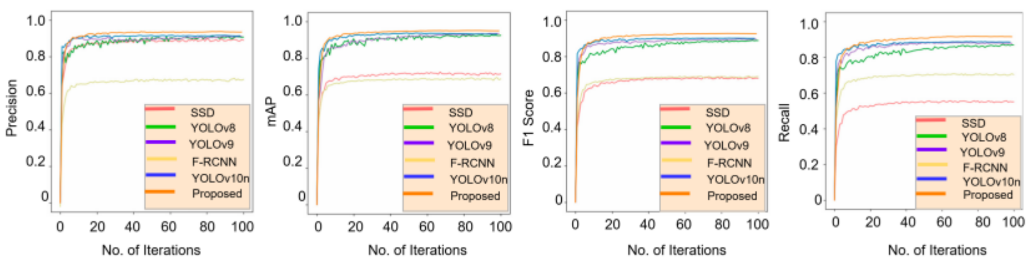

Figure 5. Proposed model training trajectories for several performance assessment metrics..

Figure 6. Implementation of segmentation on a limited set of permitted number plates and comparison between the baseline model YOLOv10 and the enhanced model.

The accuracy for segmentation is calculated as in Equation 10:

$$Acc_{Seg} = \frac{Correct_{Seg}}{Correct_{Seg} + Incorrect_{Seg}} \quad (10)$$

14

The accuracy for recognition is calculated as in Equation 11:

$$Acc_{Rec} = \frac{Correct_{Rec}}{Correct_{Rec} + Incorrect_{Rec}} \quad (11)$$

Figure 8 depicts the performance comparison of the model before and following the optimization of the neck. The findings of this research are displayed in Table 4.

Table 4. Number Plate Detection Ablation Experiment Results

Technique	Precision	Recall	mAP	Parameters
Model 1	86.5	79.6	88.0	2.02
Model 2	92.4	77.8	89.6	2.15
Model 3	89.6	82.6	91.0	2.40
Model 4	94.4	92.8	96.8	2.89
Model 5	91.3	92.0	96.0	2.29

The baseline model YOLOv10 backbone was substituted with the superior YOLOv10, attaining a precision of 94.89%, a recall of 92.79%, a mAP50 of 94.10%, and an F1 Score of 93.83%. The inclusion of the BiFPN component boosted accuracy to 93.14% and AP50 to 55.01%, although recall slightly declined to 84.90%, giving an input total of 2.25 M. A further incorporation of the SEAM component boosted recall and accuracy to 93.69% and 92.61%, as well as AP50 to 59.80%. Still, the variable quantity climbed to 2.45 million [75-76]. The introduction of an extra recognition layer greatly increased the accuracy of the model, achieving an accuracy of 93.84%, recall of 93.15%, and an AP50 of 62.58%, resulting in the parameter count expanding to 3.12 M. Figure 7 depicts an illustration of the model's accuracy before and after the integration of the minor object identification layer. Table 5 illustrates the many alterations to the model. The experiments demonstrate that the proposed improved YOLOv10 significantly enhances model detection accuracy with a little rise in computational and parametric metrics [77].

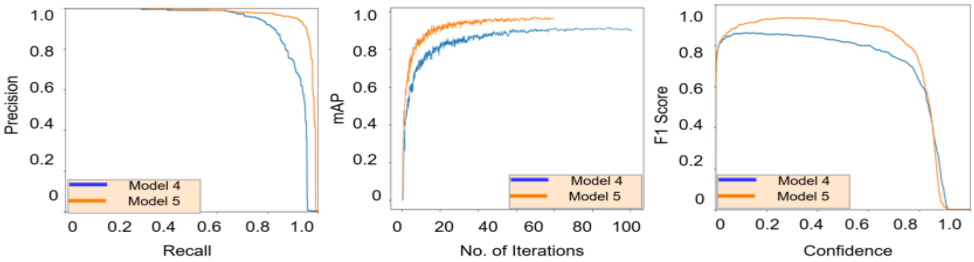

Figure 7. Assessment of the metrics with model 4 & 5

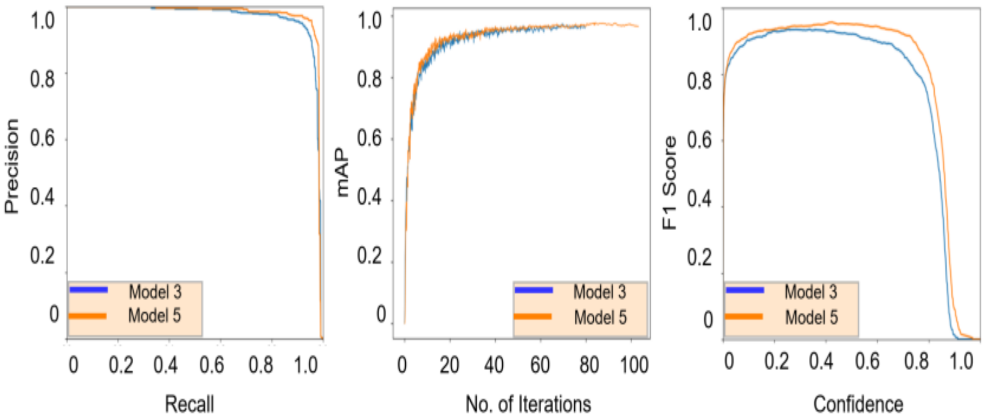

Figure 8. Assessment of the evaluation metrics with model 3 & 5

Table 5. Different Improvements Techniques Applied

Technique	Model1	Model2	Model3	Model4	Model5	Proposed
BiFPN	-	No	Yes	Yes	Yes	Yes
SEAM	-	No	Yes	No	Yes	Yes
DySample	-	Yes	Yes	Yes	Yes	Yes
GCNet	-	No	Yes	No	No	Yes
DepthConv	-	Yes	Yes	Yes	Yes	Yes

4.3. Model Performance Before and After Improvement

Table 6 demonstrates that the proposed YOLOv10 enhanced detection precision relative to YOLOv10 by 2.13, 3.20, and 1.27 percentage points for small number plates, obscured number plates, and number plates in complicated environments, respectively. The findings demonstrate that the optimized model outperformed the original YOLOv10 on the training set, attaining an 85% enhancement in mAP.

Table 6. Model mAP50 for different number plates

Licensed Number Plate	Base Model YOLOv10	Improved Model YOLOv10
Small Targets	89.10	91.23
Obscured Objects	88.21	91.45
Complex Environment	90.10	91.37

4.4. Comparative Analysis with other Recognized Classifiers

Table 7 illustrates that the proposed classifier achieved the highest accuracy rate of 97.87%, alongside an impressive computation time of only 2.1 milliseconds per image. MobileNet [76] demonstrated the quickest processing time at 1.1 milliseconds per image, although its accuracy was significantly inferior at 89.31%. This illustrates that while MobileNet analyzes images more rapidly, it significantly sacrifices accuracy, in contrast to the CNN classifier, which adeptly preserves a balance between speed and precision.

This study used YOLOv10 as the baseline model for car number plate detection. This research investigated the incorporation of diverse attention mechanisms and network architectures into the original YOLOv10 model to improve its object identification efficacy. In comparison to the SS-D, Faster R-CNN, YOLO v8, YOLO v9, and YOLOv10, the enhanced model exhibited a greater correlation and accuracy in the detection of number plate objects [78] relative to the results of the other methods. This study obtained promising results by employing the enhanced YOLOv10 for object detection, followed by a CNN classifier for licensed number plate segmentation, and ultimately character recognition on the number plate.

Table 7. Comparative Analysis with SOTA Methods on AOLP Dataset.

Technique	Accuracy (%)	Computation time (%)
Res-Net[62]	91.78	4.15
Efficient-Net[63]	93.20	13.21
Mobile-Net[64]	89.31	1.10
NAS-Net[65]	95.70	13.71
X-ception[66]	88.54	6.16
CNN[73]	92.89	5.32
C-RNN[74]	96.65	11.50
[Improved YOLOv10]	97.87	2.1

5. Conclusion

Intelligent vehicular technology has revolutionized human life in the modern era. Identifying license plates on bikes, cars, trucks, and vans aids law enforcement, surveillance, and toll collection; however, obstacles emerge due to low resolution, environmental pollution, insufficient lighting, and obstruction. Contemporary approaches can identify license plates in simple conditions. This research presents an innovative upgraded YOLOv10 for license plate detection, employing a cross-stage partial bottleneck CNN classifier for alphanumeric character recognition in a multimodal framework. This study incorporated BiFPN, SEAM, and GCNet modules into an improved YOLOv10 model. GCNet employs the global context block to record long-range relationships in the image, improving the model’s capacity to understand complex information more precisely. This study employs a comprehensive dataset augmentation strategy to enhance the recognition of partial plates under simulated adverse conditions, significantly improving the model’s robustness in real-world scenarios. The experimental findings demonstrate that this technique achieved a precision of 94.89%, a recall of 92.79%, an F1 Score of 93.83%, and a mAP of 94.10% in number plate detection, exceeding the benchmark YOLOv10 model by 3.22% in precision, 4.01% in recall, 3.83% in F1 Score, and 0.56% in mAP on the customized AOLP dataset. The R2 value is 96.15%, accompanied by a RMSE of 4.01%. The proposed model attains an accuracy of 97.87% and a processing duration of 2.1 milliseconds.

References

1. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., Ding, G.: Yolov10: Real-time end-to-end object detection. arXiv preprint arXiv:2405.14458 (2024)
2. Bogdoll, D., Nitsche, M., Z¨ollner, J.M.: Anomaly detection in autonomous driving: A survey. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4488–4499 (2022)
3. Dos Reis, D.H., Welfer, D., De Souza Leite Cuadros, M.A., Gamarra, D.F.T.: Mobile robot navigation using an object recognition software with rgbd images and the yolo algorithm. Applied Artificial Intelligence 33, 1290–1305 (2019)
4. Zeng, F., Dong, B., Zhang, Y., Wang, T., Zhang, X., Wei, Y.: Motr: End-to-end multiple-object tracking with transformer. In: European Conference on Computer Vision, pp. 659–675 (2022). Springer
5. Ettalibi, A., Elouadi, A., Mansour, A.: Ai and computer vision-based real-time quality control: A review of industrial applications. Procedia Computer Science 231, 212–220 (2024)
6. Li, J., Zhou, Z., Yang, J., Pepe, A., Gsaxner, C., Luijten, G., Qu, C., Zhang, T., Chen, X., Li, W., et al.: Medshapenet–a large-scale dataset of 3d medical shapes for computer vision. arXiv preprint arXiv:2308.16139 (2023)
7. Esser, P., Rombach, R., Ommer, B.: Taming transformers for high-resolution image synthesis. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12873–12883 (2021)
8. Ditta, A., Ahmed, M.M., Mazhar, T., Shahzad, T., Alahmed, Y., Hamam, H.: Number plate recognition smart parking management system using iot. Measurement: Sensors 37, 101409 (2025)
9. Chu, Z.: D-yolo a robust framework for object detection in adverse weather conditions. arXiv preprint arXiv:2403.09233 (2024)
10. Li, C., Yang, X., Wang, G., Zheng, A., Tan, C., Tang, J.: Disentangled generation network for enlarged license plate recognition and a unified dataset. Computer Vision and Image Understanding 238, 103880 (2024)
11. Deng, Y., Wang, G., Li, C., Wang, W., Zhang, C., Tang, J.: Collaborative license plate recognition via association enhancement network with auxiliary learning and a unified benchmark. IEEE Transactions on Multimedia (2024)
12. Tao, L., Hong, S., Lin, Y., Chen, Y., He, P., Tie, Z.: A real-time license plate detection and recognition model in unconstrained scenarios. Sensors 24, 2791 (2024)
13. Ammar, A., Koubaa, A., Boulila, W., Benjdira, B., Alhabashi, Y.: A multi-stage deep-learning-based vehicle and license plate recognition system with real-time edge inference. Sensors 23, 2120 (2023)
14. Senthil Kumar, K., Abdullah Safwan, K.: Accelerating object detection with yolov4 for real-time applications. arXiv e-prints, 2410 (2024)
15. Liu, K., Wu, P., Xia, T., Liu, Y., Guo, M., Zhe, W., Cheng, Y.: Fusion-based rear license plates detection and recognition considering enlarged prints. IEEE Transactions on Instrumentation and Measurement (2024)
16. Khan, M.M., Ilyas, M.U., Khan, I.R., Alshomrani, S.M., Rahardja, S.: License plate recognition methods employing neural networks. IEEE Access 11, 73613– 73646 (2023)
17. Yang, H., Wang, Y., Xia, Z., Guo, X., Li, Y., Malamba Onanga, A.L., Liu, M., Zhang, X.: Lprnet improvement for license plate recognition in complex environ- ments. In: Proceedings of the 2024 8th International Conference on Deep Learning Technologies (ICDLT), pp. 29–33 (2024)
18. Saad, A., Sheikh, U.U., Alyasseri, Z.A.A.: An efficient layout index characters for automatic license plate recognition system based on the yolo-v8 detector. In: 2024 IEEE 8th International Conference on Signal and Image Processing Applications (ICSIPA), pp. 1–5 (2024). IEEE
19. Han, R., Sun, S.: Research on a personalized license plate recognition algorithm based on the lprnet model. In: 2025 IEEE 17th International Conference on Computer Research and Development (ICCRD), pp. 89–95 (2025). IEEE
20. Putluru, A.R., Pannuru, V.R., AshokKumar, C., Kumar, C.J.: An optimized yolo- based license plate recognition system with integrated privacy safeguards. In: 2025 International Conference on Machine Learning and Autonomous Systems (ICMLAS), pp. 1053–1060 (2025). IEEE
21. Pattanaik, A., Balabantaray, R.C.: Enhancement of license plate recognition per- formance using xception with mish activation function. Multimedia tools and applications 82, 16793–16815 (2023)
22. Liu, Q., Chen, S.-L., Chen, Y.-X., Yin, X.-C.: Improving license plate recognition via diverse stylistic plate generation. Pattern Recognition Letters 183, 117–124 (2024)
23. Sajid, M., Malik, K. R., Khan, A. H., Iqbal, S., Alaulamie, A. A., & Ilyas, Q. M. (2025). Next-generation diabetes diagnosis and personalized diet-activity management: A hybrid ensemble paradigm. PloS one, 20(1), e0307718.
24. Alheraki, M., Al-Matham, R., Al-Khalifa, H.: Handwritten arabic character recognition for children writing using convolutional neural network and stroke identification. Human-Centric Intelligent Systems 3, 147–159 (2023)

25. Basheer Ahmed, M.I., Zaghdoud, R., Ahmed, M.S., Sendi, R., Alsharif, S., Alab- dulkarim, J., Albin Saad, B.A., Alsabt, R., Rahman, A., Krishnasamy, G.: A real-time computer vision based approach to detection and classification of traffic incidents. Big data and cognitive computing 7, 22 (2023)
26. Pradhan, G., Prusty, M.R., Negi, V.S., Chinara, S.: Advanced iot-integrated park- ing systems with automated license plate recognition and payment management. Scientific Reports 15(1), 2388 (2025)
27. Mustafa, T., Karabatak, M.: Real time car model and plate detection system by using deep learning architectures. IEEE Access (2024)
28. Davy, M.K., Banda, P.J., Hamweendo, A.: Automatic vehicle number plate recognition system. Phys. Astron Int. J 7, 69–72 (2023)
29. Sajid, M., Malik, K.R., Almogren, A. et al. Enhancing intrusion detection: a hybrid machine and deep learning approach. J Cloud Comp 13, 123 (2024).
30. Jin, M., Sun, C., Hu, Y.: An intelligent traffic detection approach for vehicles on highway using pattern recognition and deep learning. Soft Computing 27(8), 5041–5052 (2023)
31. Sankaranarayanan, M., Mala, C., Mathew, S.: Efficient vehicle detection for traf- fic video-based intelligent transportation systems applications using recurrent architecture. Multimedia Tools and Applications 82(25), 39015–39033 (2023)
32. Chopade, R., Ayarekar, B., Mangore, S., Yadav, A., Gurav, U., Patil, T., Prab- havalkar, V., Chanchal, A.K.: Automatic number plate recognition: A deep dive into yolov8 and resnet-50 integration. In: 2024 International Conference on Integrated Circuits and Communication Systems (ICICACS), pp. 1–8 (2024). IEEE
33. Moussaoui, H., Akkad, N.E., Benslimane, M., El-Shafai, W., Baihan, A., Hewage, C., Rathore, R.S.: Enhancing automated vehicle identification by integrating yolo v8 and ocr techniques for high-precision license plate detection and recognition. Scientific Reports 14(1), 14389 (2024)
34. Selmi, Z., Halima, M.B., Pal, U., Alimi, M.A.: Delp-dar system for license plate detection and recognition. Pattern Recognition Letters 129, 213–223 (2020)
35. Kaur, H., Bansal, S., Kumar, M., Mittal, A., Kumar, K.: Worddeepnet: handwrit- ten gurumukhi word recognition using convolutional neural network. Multimedia Tools and Applications 82(30), 46763–46788 (2023)
36. Kaur, H., Kumar, M., Gupta, A., Sachdeva, M., Mittal, A., Kumar, K.: Bagging: An ensemble approach for recognition of handwritten place names in gurumukhi script. ACM Transactions on Asian and Low-Resource Language Information Processing 22(7), 1–25 (2023)
37. Moussaoui, H., Benslimane, M., El Akkad, N.: Image segmentation approach based on hybridization between k-means and mask r-cnn. In: WITS 2020: Pro- ceedings of the 6th International Conference on Wireless Technologies, Embedded, and Intelligent Systems, pp. 821–830 (2021). Springer
38. Li, Y., Li, J., Meng, P.: Attention-yolov4: a real-time and high-accurate traffic sign detection algorithm. Multimedia tools and applications 82, 7567–7582 (2023)
39. Lv, W., Zhao, Y., Chang, Q., Huang, K., Wang, G., Liu, Y.: Rt-detrv2: Improved baseline with bag-of-freebies for real-time detection transformer. arXiv preprint arXiv:2407.17140 (2024)
40. Han, X., Chang, J., Wang, K.: You only look once: unified, real-time object detection. Procedia Computer Science 183, 61–72 (2021)
41. Sajid, M., Razzaq Malik, K., Ur Rehman, A., Safdar Malik, T., Alajmi, M., Haider Khan, A., ... & Hussen, S. (2025). Leveraging two-dimensional pre-trained vision transformers for three-dimensional model generation via masked autoencoders. Scientific Reports, 15(1), 3164.
42. cashier system based on faster region-based convolution neural network and yolo9000. In: 2021 17th International Conference on Quality in Research (QIR): International Symposium on Electrical and Computer Engineering, pp. 153–157 (2021). IEEE
43. Sundaresan Geetha, A., Alif, M.A.R., Hussain, M., Allen, P.: Comparative anal- ysis of yolov8 and yolov10 in vehicle detection: Performance metrics and model efficacy. Vehicles 6, 1364–1382 (2024)
44. Bochkovskiy, A., Wang, C.-Y., Liao, H.-Y.M.: Yolov4: Optimal speed and accuracy of object detection. arXiv preprint arXiv:2004.10934 (2020)
45. Jocher, G., Chaurasia, A., Stoken, A., Borovec, J., Kwon, Y., Michael, K., Fang, J., Yifu, Z., Wong, C., Montes, D., et al.: ultralytics/yolov5: v7. 0-yolov5 sota realtime instance segmentation. Zenodo (2022)
46. Wang, C.-Y., Liao, H.-Y.M., Wu, Y.-H., Chen, P.-Y., Hsieh, J.-W., Yeh, I.-H.:
47. Cspnet: A new backbone that can enhance learning capability of cnn. In: Proceed- ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pp. 390–391 (2020)

48. Noel, M.M., Ghalib, M.R., Subramaniam, P.R., Venugopal, C., et al.: Novel dark- net traffic data synthesis using generative adversarial networks enhanced with oscillatory growing cosine unit activated convolution layers. International Journal of Cognitive Computing in Engineering 5, 58–65 (2024)
49. Safaldin, M., Zaghden, N., Mejdoub, M.: An improved yolov8 to detect moving objects. IEEE Access (2024)
50. Wang, C., He, W., Nie, Y., Guo, J., Liu, C., Wang, Y., Han, K.: Gold-yolo: Effi- cient object detector via gather-and-distribute mechanism. Advances in Neural Information Processing Systems 36 (2024)
51. Sajid, M., Aslam, N., Abid, M. K., & Fuzail, M. (2022). RDED: Recommendation of Diet and Exercise for Diabetes Patients using Restricted Boltzmann Machine.
52. Su, Z., Adam, A., Nasrudin, M.F., Prabuwono, A.S.: Proposal-free fully con- volutional network: Object detection based on a box map. Sensors 24, 3529 (2024)
53. Zhang, H., Shao, F., Chu, W., Dai, J., Li, X., Zhang, X., Gong, C.: Faster r- cnn based on frame difference and spatiotemporal context for vehicle detection. Signal, Image and Video Processing 18, 7013–7027 (2024)
54. Afsar, M.M., Bakhshi, A.D., Hussain, E., Iqbal, J.: A deep learning-based frame- work for object recognition in ecological environments with dense focal loss and occlusion. Neural Computing and Applications 36, 9591–9604 (2024)
55. Lin, Z., Wang, Y., Zhang, J., Chu, X.: Dynamicdet: A unified dynamic archi- tecture for object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6282–6291 (2023)
56. Ye, M., Ke, L., Li, S., Tai, Y.-W., Tang, C.-K., Danelljan, M., Yu, F.: Cascade- detr: delving into high-quality universal object detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 6704–6714 (2023)
57. Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., Tomizuka, M., Li, L., Yuan, Z., Wang, C., et al.: Sparse r-cnn: End-to-end object detection with learnable proposals. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 14454–14463 (2021)
58. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End- to-end object detection with transformers. In: European Conference on Computer Vision, pp. 213–229 (2020). Springer
59. Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable detr: Deformable transformers for end-to-end object detection. arXiv preprint arXiv:2010.04159 (2020)
60. Sajid, M., Khan, A. H., Malik, K. R., Khan, J. A., & Alwadain, A. (2025). A new approach of anomaly detection in shopping center surveillance videos for theft prevention based on RLCNN model. PeerJ Computer Science, 11, e2944.
61. Shaw, A.K., Sarmah, P., Mishra, B., BikashSingh, T.: Iot based real-time automatic number plate detection using opencv. In: NIELIT's International Con- ference on Communication, Electronics and Digital Technologies, pp. 173–187 (2024). Springer
62. Jawale, M., William, P., Pawar, A., Marriwala, N.: Implementation of number plate detection system for vehicle registration using iot and recognition using cnn. Measurement: Sensors 27, 100761 (2023)
63. Pandey, V., Choubey, S., Patra, J., Mall, S., Choubey, A.: A novel method for indian vehicle registration number plate detection and recognition using cnn. Journal of Intelligent & Fuzzy Systems (Preprint), 1–23 (2024)
64. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recogni- tion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
65. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neu- ral networks. In: International Conference on Machine Learning, pp. 6105–6114 (2019). PMLR
66. Howard, A.G.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017)
67. Zoph, B.: Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578 (2016)
68. Chollet, F.: Xception: Deep learning with depthwise separable convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1251–1258 (2017)
69. Hsu, G.-S., Chen, J.-C., Chung, Y.-Z.: Application-oriented license plate recog- nition. IEEE transactions on vehicular technology 62, 552–561 (2012)
70. Sajid, M., Khan, A.H., Malik, T.S., Bilal, A., Ahmad, Z., Sarwar, R.: Enhancing melanoma diagnostic: Harnessing the synergy of ai and cnns for groundbreaking advances in early melanoma detection and treatment strategies. International Journal of Imaging Systems and Technology, https://doi.org/10.1002/ima.70016 35 (2024)
71. Tan, M., Pang, R., Le, Q.V.: Efficientdet: Scalable and efficient object detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10781–10790 (2020)

72. Lin, T.-Y., Doll´ar, P., Girshick, R., He, K., Hariharan, B., Belongie, S.: Feature pyramid networks for object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2117–2125 (2017)
73. Liu, S., Qi, L., Qin, H., Shi, J., Jia, J.: Path aggregation network for instance segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8759–8768 (2018)
74. Yu, Z., Huang, H., Chen, W., Su, Y., Liu, Y., Wang, X.: Yolo-facev2: A scale and occlusion aware face detector. Pattern Recognition 155, 110714 (2024)
75. Tung, C.-L., Wang, C.-H., Peng, B.-S.: A deep learning model of dual-stage license plate recognition applicable to the data processing industry. Mathematical Problems in Engineering 2021, 3723715 (2021)
76. Sajid, M., Sanaullah, M., Fuzail, M., Malik, T. S., & Shuhidan, S. M. (2025). Comparative analysis of text-based plagiarism detection techniques. PloS one, 20(4), e0319551.
77. Malik, K. R., Sajid, M., Almogren, A., Malik, T. S., Khan, A. H., Altameem, A., ... & Hussen, S. (2025). A hybrid steganography framework using DCT and GAN for secure data communication in the big data era. Scientific Reports, 15(1), 1-23.
78. Ahmad, H., Sajid, M., Mazhar, F., & Fuzail, M. (2025). Mapping Unseen Connections: Graph Clustering to Expose User Interaction Patterns. Journal of Future Artificial Intelligence and Technologies, 1(4), 474-496.