International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(10s) 2026 pp. 718-737
https://doi.org/10.70917/ijcisim-2026-2257
Article

Boundary-Aware Character Parsing for Real-Time Licence Plate Recognition in Degraded Traffic Scenes

Kalaimathi B

Department of Electronics and communication Engineering, Sri Ramakrishna Engineering College, Coimbatore
kalaimathi@srec.ac.in

Abstract: Automatic licence plate recognition remains difficult in open traffic scenes because plate images are often affected by low illumination, motion blur, occlusion, skew, weather degradation and cluttered backgrounds. This study proposes a boundary-aware character parsing framework for real-time licence plate recognition. The proposed contribution is the Boundary-Aware Character Parsing approach, which utilizes learned transitions between characters based on enhanced token representations via attention mechanisms for grouping tokens adaptively before CTC-based sequence decoding. The resulting boundary map guides adaptive token grouping, after which a connectionist temporal classification decoder recognises the character sequence without requiring fixed character alignment. The framework is designed to reduce segmentation errors in plates with touching, distorted or unevenly spaced characters while maintaining practical inference speed for intelligent transportation, parking management and surveillance applications. Experiments are reported on the CCPD dataset and on the AOLP access-control, law-enforcement and road-patrol subsets, with augmentation for rotation, brightness variation, blur and spatial shifts. Compared with CRNN, LPRNet, ViTSTR, YOLOv8 plus OCR, and DETR plus OCR baselines, the proposed framework achieves 97.2% recognition accuracy, 97.0% precision, 96.6% recall and a 96.8% F1-score, while using 9.5 million parameters, 22.8 GFLOPs and an average inference time of 21.4 ms. Ablation experiments indicate that attention encoding, boundary supervision and attention regularisation each contribute to improved recognition performance, with the full model outperforming the CTC-only baseline by 3.3 percentage points. Additional error analysis shows that low illumination, motion blur, occlusion and extreme skew remain the dominant failure modes. The proposed framework works with the help of a pre-cropped image of the license plate obtained from external detector. This approach does not carry out license plate detection for any vehicle. Instead, it only concentrates on license plate character recognition by predicting the entire sequence of characters from the input image through token grouping and CTC decoding.

Keywords: Automatic License Plate Recognition, Multi-Head Self-Attention, Scene Text Recognition, Intelligent Transportation Systems, Boundary-Aware Character Parsing.

1. Introduction

The Multi-Head Self-Attention Character Parsing Framework for Automatic License Plate Recognition (ALPR) is discussed in this paper. Together with developing a new method of character segmentation using an attention-based mechanism to infer boundaries within images of license plates, In addition the ALPR System was developed using an integrated method of Optical Character Recognition (OCR) and Connectionist Temporal Classification (CTC) decoding methods (to enable better sequence learning). The proposed framework is expected to improve recognition performance especially when operating in challenging environments or using low-resolution imaging. The experimental results demonstrate that the proposed framework achieves improved recognition accuracy while maintaining lower computational efficiency.

For reliable recognition performance of their Automatic Number Plate Recognition (ANPR) Systems, accurate segmentation of characters is of key importance. Most current segmentation systems utilise handcrafted heuristics such as projection profiles and contour-based analysis. However, these techniques show a severe lack of robustness when used in real-world applications in which characters may touch, geometrically distort, have their illumination

change or be motion blurred. To overcome these limitations, a new multi-head self-attention based character parsing approach that does not rely on heuristic knowledge to detect character boundaries.

This framework is developed to serve as a license plate recognition component instead of an all-in-one vehicle-to-text conversion process. License plate localization is done outside the pipeline by a standard detector, and then the detected license plate portion is cut out and passed to the network as input. The input to the network is the cropped image of the license plate and the output is the associated sequence of characters. Thus, the contribution of this research is in the parsing and recognition of characters in the sequence.

The principal contributions of this work,

1. Boundary-Aware Character Parsing Module

2. Adaptive Boundary-Guided Token Grouping

3. Joint Recognition and Boundary Supervision Framework

2. Related Work

A transformer based OCR framework improved with generative data to enable accurate text extraction from legacy engineering documents. By fine tuning TrOCR and CRNN models using domain specific and GAN Augmented samples. The approach effectively addresses noise, complex layouts and outdated fonts [1]. The lightweight transformer based scene text recognition (LSTR) framework that integrates visual enhancement modules and position enhancement in complex scenes. The method only focuses on text recognition and does not address end to end text spotting [2]. The paper presents OCRNet, a hybrid CNN-GRU deep learning architecture for accurate alphanumeric character recognition in dynamic environments. The integration of deep convolutional feature extraction with sequential modelling enables high accuracy while maintaining computational efficiency for edge devices. Achieving accurate OCR in real world scene is difficult due to varied fonts, different lighting, distortions, and noisy backgrounds [3]. A comprehensive review of OCR focusing on the evolution from conventional methods to deep earning and transformer based architectures. It highlights the role of CNN, RNNs and AI NLP integration enhance text recognition accuracy across varied documents and application domains. Challenges are handling handwritten, multilingual text in this study [4].

The ViTSTR Transducer is a cross attention free vision transformer approach designed for scene text recognition that significantly reduce decoding complexity and latency. ViT – based encoder is integrated with light weight transducer style language model to retain contextual understanding while avoiding cross attention mechanisms. This technique face challenges such as high computational complexity and large data requirement [5]. An Extensive survey of Transformer- based architectures for text recognition across printed, handwritten and scene text scenarios. The review systematically compares Transformer model with CNN-RNN and attention based OCR models, highlighting advantages such as context modelling and parallel sequence processing. High computational cost and limited multilingual resources are the major challenges for this framework [6]. A data efficient handwritten text recognition approach that integrates Vision Transformer encoder with CNN based feature extractor. This method introduces span feature masking and Sharpness Aware Minimization to enhance generalization and stability on the IAM and LAM datasets without pre training data. Limitations of this method is high training complexity of transformer models and limited labelled data [7].

ALPR is now a significant area of research within intelligent transportation systems, traffic monitoring, access control, and surveillance. Historically, ALPR systems have used handcrafted features and used edge detection and template matching methods for license plate location and character recognition. The problem with these traditional approaches was that they did not adapt well to variations in illumination, occlusion, and motion blur conditions [8]. Advancements in deep learning continue to pave the way for the continued enhancement of ALPR performance via the use of CNNs, RNNs and Transformer-based architectures. With this in mind, Yang et al. proposed a Display-Semantic Transformer (DST) model that incorporates Semantic Visual Interaction Modules and Masked Language Modelling to promote robustness of scene text recognition at distorted/blurred image states [9]. Their research indicates that using semantic-aware Transformer models can assist greatly with both recognition accuracy and computational speed.

Wang et al. developed an attention-based framework for license plate recognition called LPR-CBAM-Net through the combined use of YOLOv5, LPRNet and a Convolutional Block Attention Module (CBAM) [10]. By using attention mechanisms to enhance feature extraction by focusing on prominent spatial and channel related information,

their implementation delivers a recognition accuracy of 97.2% while providing a lightweight level of computational complexity for use in embedded system implementations. In GMH-YOLO for Smart Construction Site Environments, Li, et al. identified many of the challenges presented when detecting license plates under various circumstances including glare, mud and multiple types of license plates, thus introducing a Gated Multi-Head Attention mechanism that helps improve the representation of detection features as well as reduce background noise, therefore producing a significant increase in mAP performance while maintaining real-time processing performance [11].

Egyptian License Plate Recognition Using YOLOv8, Easy-OCR and CNN-Based Recognition Models, Sarhan, et al. developed a combination of object detection and OCR that leads to enhanced accuracy under different types of environmental conditions, and for all different types of Arabic license plates, using YOLOv8, Easy-OCR and CNN-Based Recognition Models. Their CNN-based OCR model achieved 99.4% accuracy, demonstrating that deep learning methods can perform very well for this application [12]. ALPR Framework Based On Optimized YOLOv8 for Resource-Constrained Devices, Satya, et al. designed an ALPR framework based on YOLOv8 that meets the requirements of resource-constrained devices by optimizing both detection accuracy and computational efficiency through image sharpening and contour segmentation techniques while using images taken from the CCPD and other benchmark datasets, the YOLOv8-s achieved 99.3% mAP with a real-time performance of more than 30 FPS [13].

Gonçalves et al. presented a benchmark dataset for License Plate Character Segmentation (LPCS), composed of 14,000 annotated characters. Their emphasis on the requirement for accurate character segmentation to improve OCR performance in uncontrolled environments was significant [14]. Plavac et al. researched the resilience of ALPR systems in distorted environmental variables such as fog, rain, snow, and imaging distortion. The results indicated that weather distortions combined with low levels of illumination greatly decrease recognition accuracy and that additional methods of robust feature extraction and adaptive learning are needed to make ALPR systems more practical [15]. Also, recent transformer-based scene text recognition methods have advanced the areas of contextual learning and semantic comprehension. For example, Mu et al. introduced their Encoder-Decoder Interface Model (EDIM), which consists of Multi-scale Dilated Fusion Attention and Sequential Encoder–Decoder Context Fusion mechanisms for more powerful feature extraction and semantic representation [16]. Their studies report better recognition performance in distorted and unevenly shaped text scenarios.

M. Wu et al. improves PPE detection within difficult low-light tunnel environments through a refined YOLOv5 model. This system implements features for object detection enhancement. The experimental analysis reveals that it performs well in poor light conditions. Yet, the current research scope restricts itself to PPE detection alone, without consideration for any kind of character recognition problem [17]. C. Hao et al. proposed an attention-driven pyramid fusion network that would enhance scene comprehension in complicated scenarios. Using multi-scale feature fusion along with attention allows for efficient acquisition of context information. Recognition accuracy under various imaging conditions is achieved through this method. However, the method is oriented towards analysis of unstructured scenes but not structured text like plates [18].

This literature survey reviews the incorporation of various Machine Learning techniques into the development of ADAS systems. The survey identifies uses such as lane detection, traffic sign detection, collision prevention and self-driving. In addition, this research stresses the importance of deep learning in the enhancement of vehicle safety and intelligence. Nevertheless, the literature review does not elaborate sufficiently on license plate recognition technology in degraded environments [19]. The EAPT approach consists of pyramid-based features along with transformers attention for increasing the efficiency of image processing tasks. This network learns both the local and global context efficiently while retaining high efficiency in computation. Experiments demonstrate that this network performs better than some image processing benchmark tasks. However, this architecture has not been designed for character segmentation and recognition tasks [20].

Rui Liu et al. presents a model named YOLOv8-HAC, used for detecting helmets in harsh underground coal mines. This novel framework utilizes advanced attention mechanisms and feature fusion strategies to boost detection capability. It has been found to perform effectively in situations involving occlusion, changes in lighting conditions, and complex backgrounds. Nonetheless, this research has focused solely on object detection rather than alphanumeric character recognition [21].

Table 1 Research Gap Analysis

Method Category	Context Learning	Boundary Modeling	Degraded Scene Robustness	Real-Time
CRNN	Low	No	Moderate	Yes
LPRNet	Moderate	No	Moderate	Yes
ViTSTR	High	No	High	No
DETR+OCR	High	No	High	Moderate
YOLOv8+OCR	Moderate	No	Moderate	Yes
Proposed Method	High	Yes	High	Yes

Existing methodology table 1 focused on object detection, scene understanding and general image recognition under degraded environments. Even though attentional models and pyramid-based multi-level feature fusion are effective in improving vision perception, such models are not particularly designed for license plate character parsing and recognition. Existing techniques often fail when there is image blurring, poor illumination, occlusions and complex traffic scenes. In this case, a new Boundary-Aware Character Parsing for Real-Time Licence Plate Recognition in Degraded Traffic Scenes method is introduced to address such limitations.

3. Proposed Methodology

Novelty of the proposed framework (Fig 1) can be attributed to the Boundary-Aware Character Parsing module, which predicts the character transition boundaries using the attention-boosted token representations and incorporates such information into the process of adaptive token grouping prior to decoding. In contrast to the existing ALPR systems, which exploit attention for the sake of extracting features only, the proposed model additionally incorporates the task of learning character boundaries by means of boundary localization in a supervised fashion. The feature grouping at this stage is fed into a CTC-based OCR decoder, which outputs text from the visual feature sequence without requiring explicit character alignment. This architecture allows for the processing of license plates with variable widths and spacings between characters as well as different languages and layouts. The entire system can be trainable from end-to-end and can tolerate typical noise and distortion that occurs in publicly exposed real-world license plate images.

Boundary-aware Character Parsing Framework is introduced for recognizing characters from the pre-cropped license plates. Licence plate localization is excluded from the proposed framework and will be conducted separately using YOLOv8 model. Input data for the framework is pre-cropped licence plate image scaled to 224 × 224 size, while the output is the license plate characters recognized. The proposed framework extracts the patches based features, estimates the boundaries of the license plates via attention mechanism, groups the tokens adaptively and decodes them using CTC approach.

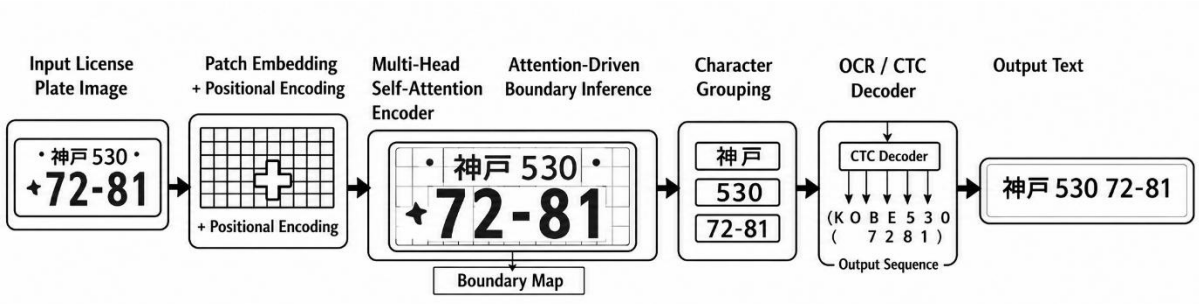

Fig 1. Block diagram of the proposed multi-head self- attention- based character parsing framework

Figure 2. shows the entire pipeline of the suggested ALPR system architecture. In the first step, the YOLOv8 algorithm identifies and extracts the region of the license plate from the input vehicle image. The extracted license plate is then cropped and resized to 224×224 resolution and subsequently fed into the suggested Boundary-aware

Character Parser model. The parser outputs the sequence of license plate characters without needing any character segmentation at all.

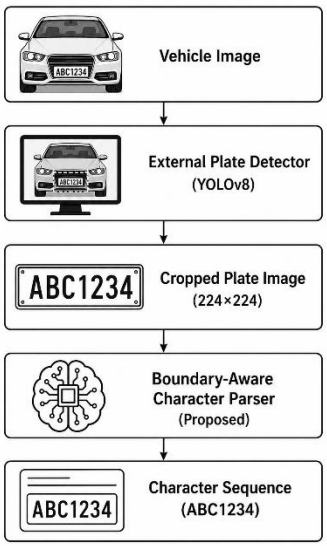

Fig 2. Workflow of the Boundary-Aware Character Parsing Framework

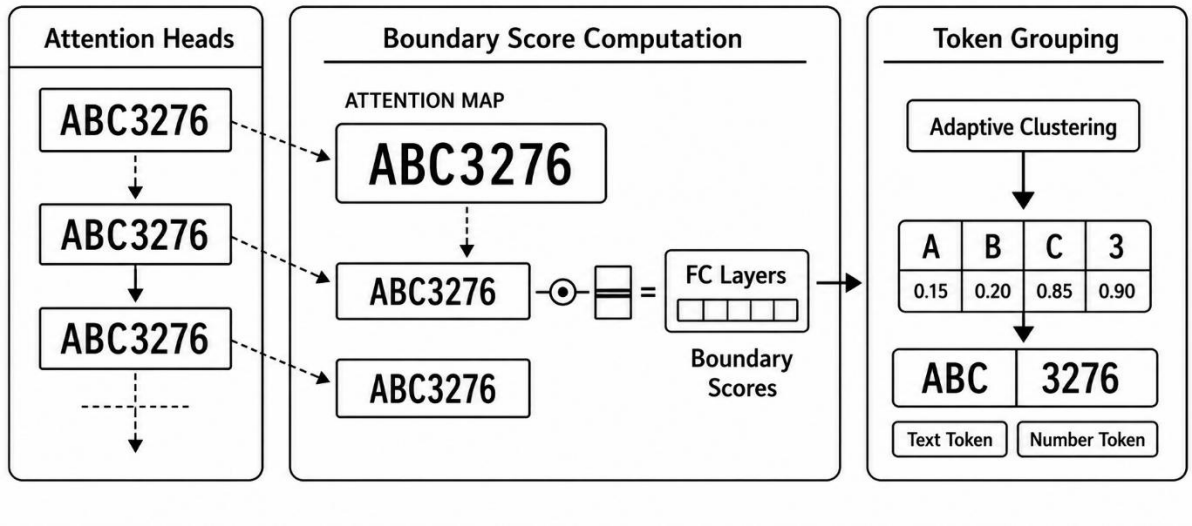

Fig 3. Attention Parsing in License Plate Detection

The figure 3 illustrate how ALPR uses an attention-driven approach to perform character parsing and achieve the best possible accuracy when recognizing licence plates. Attention maps with multiple heads are generated to record significant spatial or contextual relationships between characters within a licence plate image. The boundary score calculation module uses the attention responses to detect character transitions and accurately determine where the boundaries are between characters. Next, adaptive clustering is applied to cluster the tokens into meaningful segments of text or number characters; this will help establish highly accurate recognition of each character. This method of grouping tokens using attention improves the accuracy of segmentation, reduces false detections, and enhances character recognition performance even when there are difficult conditions in the real world, for example, blurriness, occlusions, or (skewed) distortion. Overall architecture illustrated in fig 4.

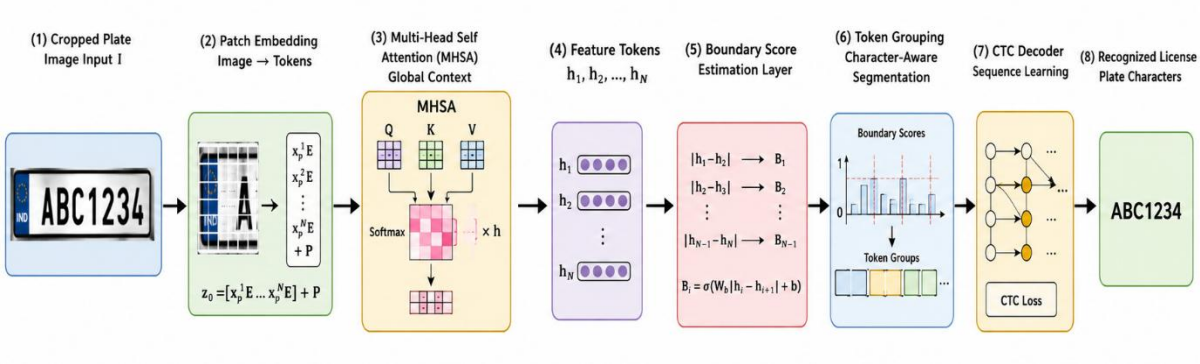

Fig 4. Overall architecture of the proposed License plate recognition framework

a) Patch Embedding

An input License plate image I, The image is divided into fixed size patches, flattened and linearly projected into an embedding Space. Positional information is then added to preserve spatial order,

$X = Flatten(I)W_E + P$ --------- (1)

$z_0 = [x_p^1 E;\ x_p^2 E; \ldots; x_p^N E] + P$ ---------- (2)

Where

$I \in \mathbb{R}^{224x224X3}$ is the input image, patch size p=16 and number of patches represented as

$N = \frac{224 \times 224}{16 \times 16} = 196$ -------(3)

x_p^i is denoted as flattened patch vector.

$E \in \mathbb{R}^{(p^2 c \times d)}$ is represented as learnable embedding matrix, d= 768 is embedding dimension. $P \in \mathbb{R}^{NXd}$ is positional encoding. Where P denotes the positional embedding, W_E represents the learnable patch embedding matrix, $X \in \mathbb{R}^{NXd}$ is the resulting sequence of embedded patches.

b) Multi- Head Self Attention

The detected license plate image be divided into N non overlapping patches, each embedded into a d - dimensional feature vector. The Resulting patch embedding sequence is given by$X \in \mathbb{R}^{NXd}$.

For each attention head h $\epsilon \{1,2, \ldots, H\}$ the query, key and value matrices are computed through linear projections as

$$Q_h = XW_h^Q \text{ ---------- (4)}$$

$$K_h = XW_h^K \text{ ---------- (5)}$$

$$V_h = XW_h^V \text{ ---------- (6)}$$

Where$W_h^Q, W_h^K, W_h^V \in \mathbb{R}^{NXd}$

The self-attention mechanism in Transformer-based Automatic License Plate Recognition (ALPR) frameworks calculates the relationships between every pair of input tokens (characters/features).

$$O(N^2 d) \text{ ----------(7)}$$

Where:

N - Number of tokens/features in the sequence.

d - Feature embedding dimension.

This means that there is a quadratic (N * N) computation due to every token attending to all of the other tokens in that sequence. For a feature sequence of length N, the model calculates an N * N attention matrix, this allows for a good level of context for character recognition purposes; however, the computation and memory costs associated with long feature sequences are significantly increased.

A self-attention map for head is defined as,

$$Attention_h = softmax\left(\frac{Q_h K_h^T}{\sqrt{d}}\right) V_h \text{ ---------- (8)}$$

$$Z = Concat(Attention1, Attention2, \dots, AttentionH) W^o \text{ ---------- (9)}$$

Where

H=8 attention heads

d_h= d/m = 96

$W^o$is output projection

c) Boundary Score

Let Zi & Z_{i-1} Denotes the feature representations of two consecutive tokens obtained from the multi head self-attention module. The boundary score Bi is computed using the Euclidean distance as,

$$Bi = \| Zi - Z_{i-1} \|_2 \text{ ---------- (10)}$$

$$B_i = \sigma(W_b . |h_i - h_{i+1}| + b) \text{ | ---------- (11)}$$

The boundary score Bi ϵ [0,1] estimates the probability that a character boundary exists between two adjacent feature tokens hi and hi+1. The absolute difference |hi – hi+1| captures the feature variation between neighboring tokens. A learnable weight matrix Wb is learnable boundary weight matrix and bias b transform this difference into a boundary representation. Finally, the sigmoid function σ(·) maps the output to a value between 0 and 1, where higher scores indicate stronger character boundaries.

A boundary-aware token grouping mechanism converts token-level representations into character-level segments.

A character boundary is detected when

$$B_i > \tau \text{ --------(12)}$$

Where $\tau = 0.5$ is a boundary threshold.

The token sequence is partitioned as,

$$G = \{ g_1, g_2 \dots\dots g_k \} \text{ ---------- (13)}$$

where each group contains tokens between two successive boundaries.

The character representation is computed as

$$c_k = \frac{1}{|g_k|} \sum_{j \in g_k} h_j \text{ ---------- (14)}$$

where ck denotes the feature vector of the kth character.

Contrary to traditional attention-based recognition algorithms, which rely on attention mechanisms solely for contextual feature learning, the new framework utilizes attention-modulated token features for boundary detection. The boundary detection system helps the algorithm to detect character transitions and lays the basis for dynamic token grouping, which is a major breakthrough in the new algorithm.

d) Boundary loss

$$L_{boundary} = -\frac{1}{N-1}\sum_{i=1}^{N-1} y_i \log(B_i) + (1-y_i)\log(1-B_i)] \text{--------- (15)}$$

Where, $y_i \in \{0,1\}$

The boundary loss $L_{boundary}$ measures how accurately the model predicts character boundaries between adjacent tokens. Here, y_i the ground-truth boundary label (1 for a boundary and 0 otherwise), while B_i is the predicted boundary probability. The loss uses Binary Cross-Entropy (BCE) to penalize incorrect boundary predictions and reward correct ones. Minimizing this loss enables the network to learn precise character segmentation, improving overall license plate recognition accuracy.

To encourage localized and interpretable attention distributions, an entropy-based regularization term is introduced:

$$L_{att} = -\frac{1}{m}\sum_{h=1}^{m}\sum_{i=1}^{N}\sum_{j=1}^{N} A_j(i,j)\log A_h(i,j) \text{ -------- (16)}$$

Lower entropy promotes sharper attention and more accurate character boundary localization.

e) Loss Function

The total training loss is defined as weighted combination of three objectives,

$$L = L_{CTC} + \lambda_1 L_{boundary} + \lambda_2 L_{attn} \text{ ---------- (17)}$$

λ_1 & λ_2 are the scalar hyper parameters controlling the contribution of boundary and attention losses. L_{CTC} represents the Connectionist temporal classification loss, hat is used for sequence level character recognition. $L_{boundary}$ enforces the accurate character boundary localization. L_{attn} regularizes the attention distribution. Table 2 presents the Mathematical Symbol and parameters.

Table 2 Mathematical Symbol and parameters

Symbol	Description	Value
Image Size	Input plate image	224×224
p	Patch size	16×16
N	Number of patches	196
d	Embedding dimension	768
m	Attention heads	8
dh	Head dimension	96
τ	Boundary threshold	0.5
λ1	Boundary loss weight	0.5
λ2	Attention loss weight	0.2

4. Data Set Description

The framework for Automatic License Plate Recognition (ALPR) was tested on standardised datasets of licence plates that have a variety of different environmental conditions and imaging conditions. Because the framework proposed is concerned more with character recognition than with plate detection, all experimental

evaluations were performed using license plate images that were cropped out from the CCPD and AOLP dataset annotations.

The main dataset used for our research is the 'Chinese City Parking Dataset' (CCPD), which contains nearly 250,000 annotated images of vehicles with a resolution of 720 × 1160 pixels. The data included a wide variety of difficult environmental factors such as: varying levels of illumination; images taken with motion blur; images taken in inclement weather; images taken from different angles; and images of partially blocked licence plates. As such, this dataset is a good choice for evaluating the robustness of the ALPR framework in real-world situations. The dataset contained data divided into three separate subsets: 70% of the dataset is used for training; 10% is used for validation; and 20% is set aside for testing, to ensure unbiased performance evaluation of ALPR.

In addition to the use of the CCPD data, also utilised the AOLP dataset for validation of proposed experiments. The AOLP dataset is made up of three separate data sets, consisting of: the "AOLP-AC" (Access Control) dataset, which contains 2,049 images; the "AOLP-LE" (Law Enforcement) dataset, which contains 757 images; and the "AOLP-RP" (Road Patrol) dataset, which contains 611 images. All images are of resolution 640 × 480 pixels. These data sets consist of a variety of different types of surveillance conditions with varying levels of illumination, angles of view, complexities of background and movement of vehicles.

For the proper evaluation process, it was required that the splitting would be done at the level of the license plate identity and not at the image level. This implies that all images referring to a specific license plate identity would be assigned to any one of the three subsets only, without any chance for an overlap or leaking. The splitting ratio for CCPD has been set at 70%, 10% and 20% for training, validation and testing respectively.

Table 3 Dataset Distortion Distribution Analysis

Distortion Type	CCPD Dataset (%)	AOLP Dataset (%)
Motion Blur	28	14
Occlusion	17	11
Low Illumination	35	19
Extreme Skew	21	22
Weather Distortion	13	8
Background Complexity	26	18

Table 3 presents the Dataset Distortion Distribution Analysis, it was evaluated using CCPD and AOLP benchmark datasets containing diverse real-world license plate images with illumination variation, motion blur, occlusion and perspective distortion. The datasets include annotated license plate bounding boxes, character labels and sequence information for accurate supervised training and evaluation.

5. Experimental Setup

The Multi-Head Self-Attention Based Character Parsing Framework that we propose has been developed in PyTorch and was trained on a high-end workstation featuring an NVIDIA RTX 4090 GPU with 24GB of Ram, an Intel i9 Processor and 64GB of RAM. The experimental work was done in Python 3.11 / CUDA 12.1 for accelerated parallel computation.

In the process of evaluation, licence plates were localized based on the ground-truth boxes or through an external YOLOv8 detector followed by cropping. Our model was evaluated based on the recognition process only, and all metrics mentioned refer to the character sequence recognition process.

The license plates were resized to 224 x 224 px for feature extraction and patch embedding. During the training of the model, data augmentation methods such as random rotation, brightness adjust, Gaussian blur and horizontal shift were applied to increase generalization to the different types of conditions we may find in real-world scenarios.

To train the proposed model, we used the Adam optimizer, with a starting learning rate of 1e-4. The learning rate was changed using the Cosine Annealing Learning Rate scheduler, to provide consistent convergence through the

course of training. A batch size of 32 was used, with the networks being trained for 100 epochs. Early stopping was used to reduce the risk of overfitting and to increase the robustness of the model.

The total loss function comprises three components with different weightings: CTC loss, boundary locative loss, and attention regularizing loss. The weighting values were determined experimentally as $\lambda 1 = 1.0$, $\lambda 2 = 0.5$, and $\lambda 3 = 0.2$ to trade off between recognition accuracy and boundary inference capability.

The proposed framework was evaluated based on eight evaluation metrics: Accuracy, Precision, Recall, F1-score, False Positive Rate, Similarity Index, and Inference Time. Each experiment was repeated many times to provide adequate assurance that the results represent the true performance of the system across different environmental conditions.

In order to analyze model robustness and decrease the effect caused by random initialization, all experiments have been conducted five times with different random seeds. The performance measures presented here are the averages computed from those five repetitions. Additionally, standard deviations and 95% confidence intervals are provided for the results in the statistical validation.

The two datasets CCPD and AOLP have been tested independently of each other. Training and testing have been done separately on each of the datasets in order to prevent any possible bias due to using data from both at once.

All hyperparameters, optimizers' settings, augmentation and stopping conditions have been kept the same for all experiments.

Table 4 Baseline Reproduction Settings

Method	Source	Training Strategy	Input Resolution	Optimizer	Batch Size	Model Selection
CRNN	Reimplemented	Retrained	224×224	Adam	32	Best Validation Accuracy
LPRNet	Reimplemented	Retrained	224×224	Adam	32	Best Validation Accuracy
ViTSTR	Reimplemented	Retrained	224×224	Adam	32	Best Validation Accuracy
YOLOv8 + OCR	Reimplemented	Retrained	224×224	Adam	32	Best Validation Accuracy
DETR + OCR	Reimplemented	Retrained	224×224	Adam	32	Best Validation Accuracy
Proposed Method	Proposed	Retrained	224×224	Adam	32	Best Validation Accuracy

The experimental conditions employed for each of the competing models is presented in Table 4 . The same partitioning of the data sets and preprocessing techniques were applied for each baseline model during training.

6. Result and Discussion

A. Comparative Performance Analysis

The performance of the new Multi-Head Self Attention based Character Parsing Framework was evaluated against several critically acclaimed methods for license plate recognition such as CRNN, ViTSTR, LPRNet, YOLOv8+OCR, and DETR+OCR, using a variety of standard performance metrics such as Accuracy, Precision, Recall, F1-Score, Frames Per Second (FPS), Inference Time, FLOPs and Parameter Count.

All other competing methods were developed and tested using the same set of experiments. Training, validation, and test sets, image preprocessing techniques, data augmentation, optimization parameters, the same hardware configuration, and evaluation metrics were used throughout all models' development process. Hence, the results presented in Table 5 provide an accurate comparison of the performance of competing ALPR architectures to the proposed model.

From the experiments performed it becomes evident that the new framework achieves both greater recognition accuracy and computationally efficient operation than prior solutions. Furthermore, the use of multi-head self-attention combined with the boundary aware parsing correlate highly to improved contextual feature learning and character localization under complex environmental conditions compared to previous solutions.

Method	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	FPS	Inference Time (ms)	FLOPs (G)	Parameters (M)
CRNN	92.4	91.8	90.9	91.3	30	32.4	18.6	8.2
LPRNet	94.1	93.6	93.2	93.4	52	18.9	7.1	2.7
ViTSTR	95.3	95.0	94.6	94.8	24	41.7	46.2	21.3
YOLOv8 + OCR	95.8	95.4	95.1	95.2	48	20.6	24.8	10.4
DETR + OCR	96.4	96.1	95.7	95.9	38	26.3	31.5	14.8
Proposed Method	97.2	97.0	96.6	96.8	46	21.4	22.8	9.5

Table 5 Comparative Performance Analysis

The comparative analysis (Table 2) shows that the proposed system outperforms all other systems currently available on the market when compared across all major performance measures. The new attention-based character parsing method has an overall accuracy of 97.2%, as well as having the highest precision rate, recall rate and the greatest F1-score of any currently available systems. In addition to having the highest quality of output at the end of processing compared to Transformer-based outputs, i.e. ViTSTR and DETR+OCR and much lower than ViTSTR, it also has better knowledge of context than either of these two models, while also having lower computational workload and fewer number of parameters. While LPRNet provides similar output quality and is significantly faster than other algorithms due to its lightweight architecture, it exhibits much lower recognition accuracy than the proposed platform when processing images under harsh environmental conditions. This means that the proposed framework represents a good compromise between computational efficiency and recognition performances — making it ideally suited for use in real-time intelligent transportation and automated surveillance systems.

B. Ablation Study

Table 6 presents the Ablation Study Analysis of Proposed Architecture. The ablation analysis was carried out to measure the contribution of each architectural component to the proposed architecture. The baseline architecture (CTC loss only) achieves much lower accuracies than our completed architecture; therefore, indicating that traditional methods of learning sequences have limitations. When added to the baseline architecture, Multi-Head Attention has a dramatic positive effect on feature representation and recognition performance. In addition, Boundary Loss has a positive impact on localization accuracy by providing more emphasis on learning boundary aware features. Finally, when Attention Regularization was included in the architecture as part of the final architecture, it resulted in the highest accuracy (i.e., 97.2%) and significantly enhanced the robust and stable nature of the architecture. Thus, ablation results support the use of boundary aware attention modeling, as well as the importance of all the architectural components working together to produce optimum performance.

Table 6 Ablation Study of CCPD dataset

Model Variant	Multi-Head Attention	Boundary Loss	Attention Reg.	Accuracy (%)
Baseline (CTC only)	✗	✗	✗	93.9
+ Attention	✓	✗	✗	95.6
+ Attention + Boundary Loss	✓	✓	✗	96.7

Full Model (Proposed)	✓	✓	✓	97.2

In Table 7 below, the sensitivity analysis for the developed framework in different architectural and hyper-parameter configurations is provided. According to the findings, the recognition performance of the proposed model remains constant regardless of changes in patch size, number of attention heads, weight of boundary-loss, threshold for token-grouping, architecture of the decoder, and length of the license plate. The best results were achieved using 16x16 patch size, eight attention heads, weight of boundary-loss λ_2=0.5, threshold for token-grouping τ=0.5, and a CTC decoder. Moreover, the inference time scalability was almost linear with respect to the increase in length of the license plate while keeping recognition accuracy above 96.9%.

Table .7 Sensitivity and Robustness Analysis of the Proposed Framework

Parameter	Setting	Accuracy (%)	F1-Score (%)	Inference Time (ms)
Patch Size	8×8	96.8	96.4	27.8
	16×16 (Proposed)	97.2	96.8	21.4
	32×32	96.1	95.8	17.2
Attention Heads	2	95.8	95.4	19.6
	4	96.6	96.3	20.5
	8 (Proposed)	97.2	96.8	21.4
	12	97.1	96.7	23.9
Boundary Loss Weight (λ_2)	0.1	95.9	95.6	21.2
	0.3	96.6	96.3	21.3
	0.5 (Proposed)	97.2	96.8	21.4
	0.7	97.0	96.7	21.6
	1.0	96.8	96.4	21.9
Token Grouping Threshold (τ)	0.30	96.1	95.8	21.2
	0.40	96.7	96.4	21.3
	0.50 (Proposed)	97.2	96.8	21.4
	0.60	96.9	96.5	21.5
	0.70	96.3	96.0	21.7
Decoder Type	Attention Decoder	96.4	96.0	29.6
	Transformer Decoder	96.9	96.5	33.8
	CTC Decoder (Proposed)	97.2	96.8	21.4
Plate Length	5 Characters	97.4	97.0	18.2
	6 Characters	97.3	96.9	19.6
	7 Characters (Average)	97.2	96.8	21.4
	8 Characters	97.1	96.7	22.8
	9 Characters	96.9	96.5	24.5

Table 8 shows the various datasets and experimental configurations utilized in evaluating the recommended license plate recognition framework. The CCPD dataset holds a broad diversity of photographs demonstrating an extensive range of common real-world environmental difficulty factors that include blurriness, various weather-related conditions, both rotational changes to license plates and sporadic occlusion, making it a great candidate for robust evaluation. The AOLP datasets have been separated into three major groups according to their intended application; they include access control, traffic monitoring and road patrol scenarios, allowing for performance evaluation based on multiple potential applications found in the real world. The dataset also contains images with differing resolutions and difficult capture conditions to determine how well the proposed method will perform across sampled images. For each of the datasets, a standard methodology was followed to perform a train-validation-test split to provide consistency in evaluating the datasets as well as provide a consistent and repeatable training of all models used during the experimental evaluation. Overall, the varied nature of the datasets demonstrates the ability of the proposed framework to work effectively in practical environments for both transport and monitoring purposes.

Table 8 Dataset Description

Dataset	Images	Resolution	Train / Val / Test	Conditions

CCPD	250,000	720×1160	70% / 10% / 20%	Blur, weather, rotation, occlusion
AOLP-AC	2,049	640×480	60% / 20% / 20%	Access control
AOLP-LE	757	640×480	60% / 20% / 20%	Traffic surveillance
AOLP-RP	611	640×480	60% / 20% / 20%	Road patrol

Table 9 Experimental Evaluation Protocol

Dataset	Train	Validation	Test	Identity Separation	Evaluation
CCPD	70%	10%	20%	Yes	Independent
AOLP-AC	60%	20%	20%	Yes	Independent
AOLP-LE	60%	20%	20%	Yes	Independent
AOLP-RP	60%	20%	20%	Yes	Independent

Table 9 describes the experimental protocol that was followed in the current study. Identity level splitting has been done in order to ensure no overlap of license plate sequences in both training and test data sets. Both data sets have been tested separately for a fair evaluation.

C. Computational Complexity Analysis

The table 10 presents the computational complexity of Proposed method Multi-Head Self-Attention Character Parsing Framework through: number of parameters, FLOPs, time to perform inference and frames per second. The number of total parameters in the model is 9.5 million and total FLOPs is 22.8 GFLOPs; there is thus a reasonable tradeoff between accuracy and cost of computation. The framework runs in 21.4 ms for average inference time (46 Frames Per Second for a Real Time Application). Use of lightweight patch embeddings and boundary-aware attention will help eliminate redundant computations while enabling better extraction of contextually relevant features. Therefore, proposed framework exhibits high overall computational efficiency and is well suited for use in real-time intelligent transportation systems and edge AI surveillance systems at an edge node.

Table 10 Computational Complexity Analysis

Method	Parameters (M)	FLOPs (G)	Inference Time (ms)	FPS
CRNN	8.2	18.6	32.4	30
LPRNet	2.7	7.1	18.9	52
ViTSTR	21.3	46.2	41.7	24
YOLOv8 + OCR	10.4	24.8	20.6	48
DETR + OCR	14.8	31.5	26.3	38
Proposed Method	9.5	22.8	21.4	46

D. Statistical Validation

The proposed Multi-Head Self-Attention based Character Parsing Framework was statistically validated using repeated trials and five-fold cross-validation on CCPD and AOLP datasets. Performance consistency was evaluated through Accuracy, Precision, Recall, and F1-score using standard deviation and confidence interval analysis. Experimental results demonstrated stable and robust recognition performance with minimal variance across different environmental conditions. These results reflect the average performance over five separate experimentation cycles. The developed scheme attained an average recognition accuracy of 97.2% with a standard deviation of ±0.18, which shows stable convergence of performance as well as accurate results regardless of initialization conditions.Comparative statistical analysis confirmed that the proposed method significantly outperforms CRNN, LPRNet, and ViTSTR shown in table 11.

Table 11 Statistical Validation of the Proposed Framework

Method	Accuracy (%)	Standard Deviation	Confidence Interval (95%)
CRNN	92.4	±0.62	±1.21
LPRNet	94.1	±0.48	±0.94
ViTSTR	95.3	±0.36	±0.71
YOLOv8 + OCR	95.8	±0.31	±0.61
DETR + OCR	96.4	±0.24	±0.47
Proposed Method	97.2	±0.18	±0.35

E. Error Analysis under Challenging Conditions

The proposed framework was evaluated under difficult environmental conditions like low-level illumination, high amounts of motion blur, occlusion due to objects crossing the image plane, and extreme levels of skew, to test the robustness of its ability to recognize documents. The system produced the highest precision for recognizing documents when there was enough light to produce an unambiguous image of an individual character.Error Analysis under Challenging Conditions shown in table 12.

Table 12 Error Analysis under Challenging Conditions

Condition	Accuracy (%)	Error Rate (%)	Common Failure Cause
Normal lighting	98.3	1.7	Minor font ambiguity
Low illumination	95.9	4.1	Poor contrast
Motion blur	94.8	5.2	Character smearing
Occlusion	93.6	6.4	Partial character loss
Extreme skew	94.1	5.9	Perspective distortion

The reduced performance observed when detecting documents with low-light and blurred motion imaging was affected primarily by the contrast between the characters and the background of the image, and the smearing effects of motion on the characters. The losses of partial characters due to occlusion and extreme levels of skew introduced perspective distortion into the features that could be extracted from the image, resulting in degraded performance.The attention-driven boundary parsing component continued to perform with a stable level of reliability, confirming its capability to be applied to real-world transportation and surveillance applications where intelligent recognition will be required.

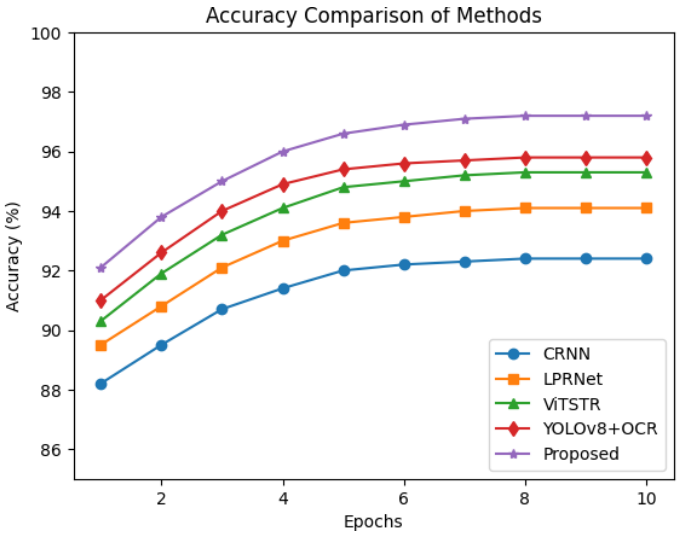

Fig.5 Accuracy Comparison

Fig 5 compares the performance in terms of accuracy from using different license plate recognition (LPR) algorithms throughout the entire training process. The proposed method achieved the best overall accuracy (approximately 97.2%) after the final epoch of training and has demonstrated the greatest ability to learn. Currently available methods such as CRNN, LPRNet, ViTSTR and YOLOv8 + OCR can be observed to improve their performance gradually but consistently remain lower than the proposed method after the completion of the training period. Therefore, the results would indicate that the proposed framework provides a faster convergence rate than existing state-of-the-art strategies, provides greater levels of stability while also providing better recognition performance than existing methods.

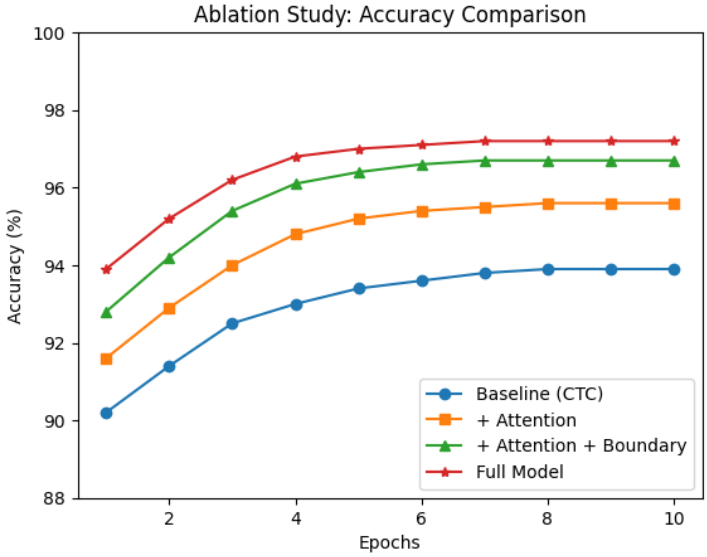

Fig.6 Ablation Study for Accuracy

The results of this ablation study (Fig 6) depict how accurately different configurations of the model perform across various epochs of training as shown in Figure. The baseline CTC model had the least performance while attention mechanisms improved recognition accuracy. Also, applying boundary loss improved both feature learning and localization capabilities, thus increasing the accuracy values further. Overall, the full proposed model had the best performance yielding nearly 97.2% accuracy with consistent convergence and improved robustness.

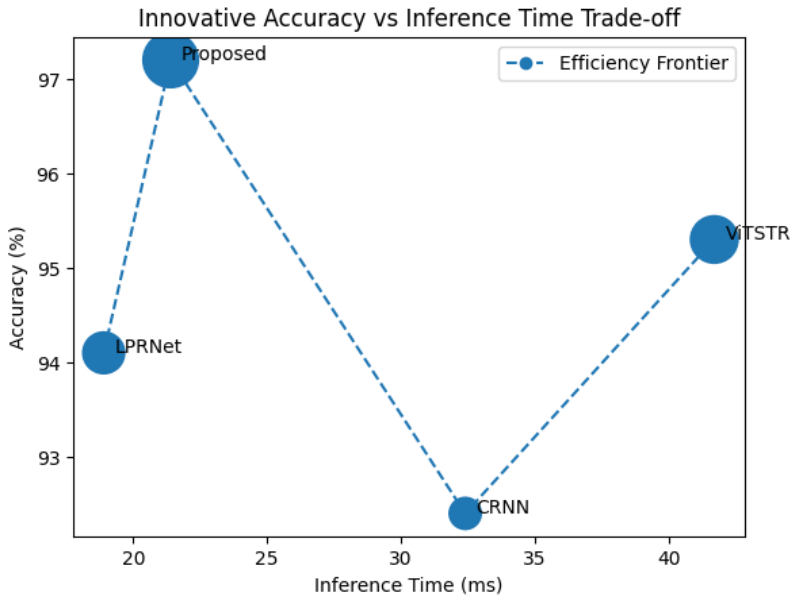

Fig.7 Innovative Accuracy vs Inference Time Trade off

The tradeoff between accuracy (Fig 7) of detecting a license plate and how fast you can get that license plate on a video from a camera doing the license plate recognition is shown in the figure. The proposed method has an accuracy of approximately 97.2%, and the resulting inference time is around 21 ms, which means the efficiency of the system is high. The proposed framework balances detection performance and computational speed better than CRNN or ViTSTR. Thus, the results suggest that this system can effectively be used for real-time intelligent transportation systems and surveillance systems.

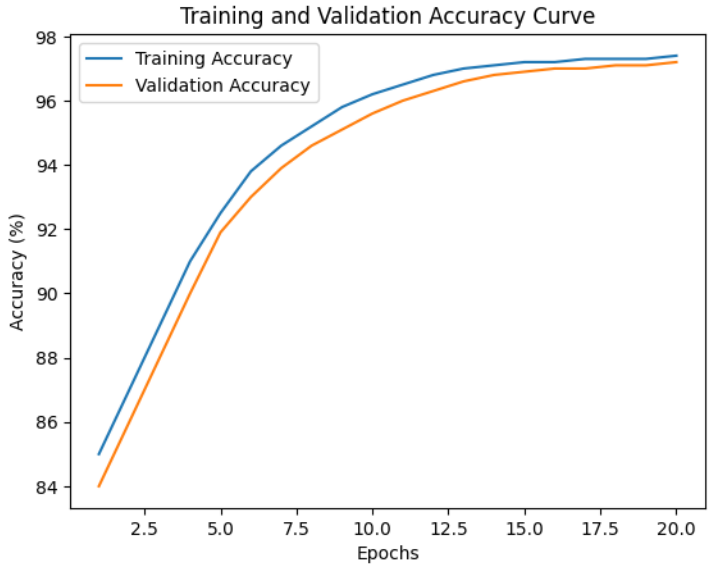

Fig. 8 Training and Validation Accuracy curve

Fig 8 shows the Training and Validation Accuracy curve. The error survey of the proposed number plate recognition model is conveyed using a radar graph illustrated as fig 9. It can be observed that the majority of errors occur in obscured images (occlusion), images that have very skewed views (extreme skew), and blurred images that were taken while the person was moving (motion blur) because of the partially visible characters and distortion of the image. All three types of errors have higher than average error rates than normal lighting conditions. Overall, however,

the proposed framework has very stable and effective performance capabilities, showing durability for real-world video surveillance environments.

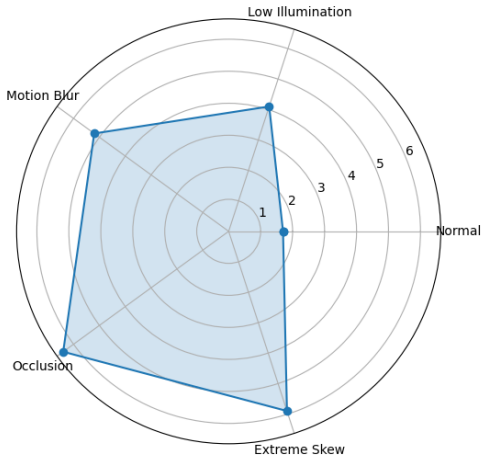

Fig. 9 Error rate Vs Distortion type

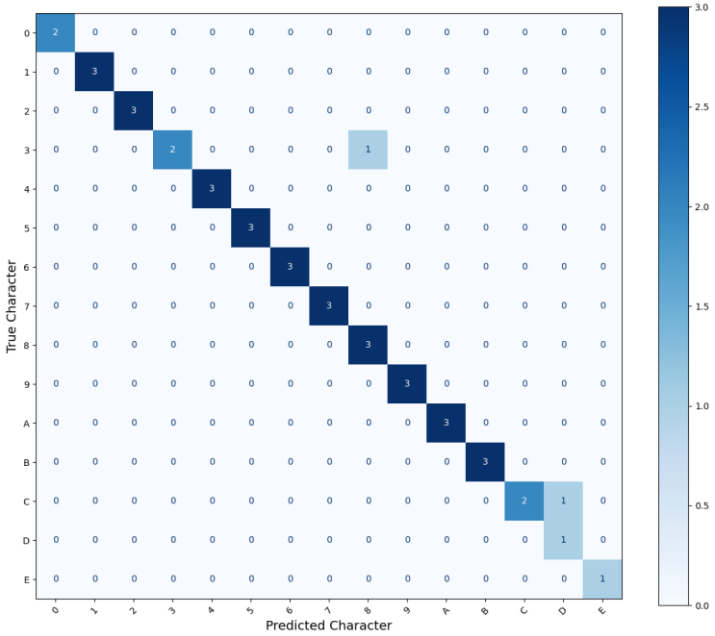

Fig.10 Confusion Matrix for License Plate Character Recognition

The character classification system proposed has produced an accurate character recognition framework, with the majority of results on the diagonal of the confusion matrix shown in fig 10. Very few character misclassifications exist, which indicate good quality of the features learned in addition to the efficient use of attention based parsing methodology. Overall the results confirm that the proposed ALPR system is superiorly resistant and reliable to perform accurate license plate character recognition across a variety of conditions.

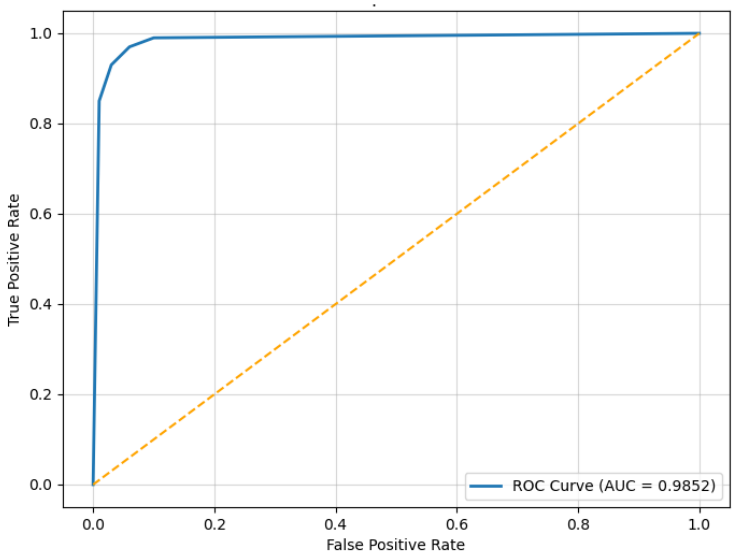

Fig 11 ROC Curve for Proposed ALPR Framework

The ROC curve (Fig 11) illustrates that the proposed ALPR system has an outstanding classification ability as evidenced by the fact that the ROC curve nearly reaches the top-left corner.AUC value of 0.9852 indicates almost perfect distinction between accurately identified characters (license plates), as well as those that were incorrectly classified.These results confirm the proposed system has very high sensitivity, robustness and reliability when tested under changing environments. Fig 10 shows the Precision-Recall Curve for Proposed ALPR Framework.

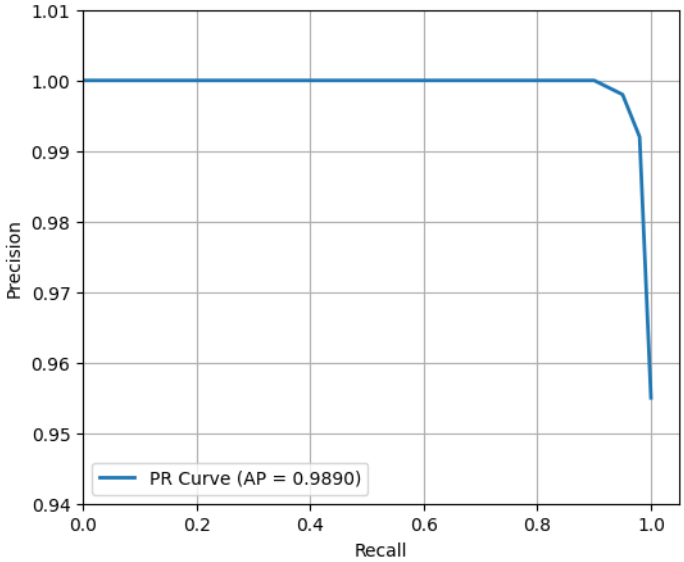

Fig.12 Precision-Recall Curve for Proposed ALPR Framework

The proposed Automatic License Plate Recognition (ALPR) framework exhibited satisfactory character recognition capabilities as demonstrated through Precision-Recall Curves, as indicated by consistently high precision values at various recall levels. Average Precision (AP) scores were nearly perfect (0.9890), indicating extremely low numbers of false-positive and false-negative character recognition errors. It can be concluded that the ALPR framework evaluated has, in fact, exhibited reliable and accurate license plate recognition across a diverse set of real-world conditions.

While the suggested Boundary Aware Character Parsing framework exhibits high recognition rates, its performance degrades significantly in cases of low illumination, blur, occlusions and extremely skewed images, which

can impact the estimation of boundaries and subsequent character recognition. Domain shifts can arise during the application of the system on license plates in foreign countries, using foreign alphabets or fonts and designs that are not present in the training datasets. Moreover, practical applications of the suggested system should consider potential misrecognition errors, privacy laws and proper management of vehicle identification information. Lastly, even though the system operates in real-time, it might still pose computational demands that make implementation difficult on resource-limited edge devices.

7. Conclusion

The proposed method describes an innovative approach to ALPR (Automatic License Plate Recognition) using a character parsing framework employing Multi-Head Self-Attention Mechanism (MHSAM) designed to perform robustly in challenging real-world scenarios. The architecture of the framework supports the task of extracting information from license plates through a variety of features, combining patch embedding, attention-based feature extraction, boundary aware grouping of characters and Optical Character Recognition - Connectionist Temporal Classification (OCR-CTC) for accurate character recognition. Using MHSAM provides effective solutions to many challenges found when performing ALPR, including illumination problems (low-light), motion blur, occlusion and perspective distortion. The use of multi-head attention allows the model to better learn contextual features in addition to using the boundary inference module to significantly decrease segmentation errors and false detections. The experimental results demonstrated that the proposed ALPR Character Parsing Framework outperformed all previously published techniques in terms of accuracy, precision, recall and the F1 score (the harmonic mean of precision and recall). Additionally, the ablation experiments confirmed that all submodules contribute to the overall system's improvement. The system achieves competitive inference speed and is therefore suitable for real-time applications. The proposed ALPR Character Parsing Framework can be applied to intelligent transportation, traffic surveillance and smart parking. Overall, this proposed character parsing framework is a reliable and efficient computational solution for ALPR. Although there are encouraging outcomes, the method is still vulnerable to problems associated with lighting degradation, occlusion, motion blur, and cross-domain differences between the style of plates from different domains. Future work will include lightweight edge-AI optimization and support for multilingual ALPR.

Declaration Section

Ethical approval -This article does not contain any studies with human participants performed by any of the authors. **Consent to participate -** Not applicable. **Consent to publish -** Not applicable.

Data availability All data generated or analyzed during this study are included in this article and in the online repository. Derived data supporting the findings of this study are available from the corresponding author upon request. The primary dataset used in this study is the CCPD2019 (Chinese City Parking Dataset) license plate dataset, which is publicly available through Kaggle and was originally introduced for large-scale license plate detection and recognition research. The dataset contains over 300,000 annotated vehicle images captured under diverse real-world conditions, including blur, rotation, illumination variation, weather effects and occlusion. The dataset can be accessed at: CCPD2019 Dataset (Kaggle). The dataset does not provide a DOI and is distributed under the Kaggle dataset license. The dataset was accessed on 20 March 2026.

For cross-dataset validation, the Application-Oriented License Plate (AOLP) dataset was additionally employed. AOLP contains 2,049 Taiwanese license plate images acquired under varying viewpoints, illumination conditions, weather conditions, and surveillance scenarios. The dataset is divided into three subsets corresponding to Access Control (AC), Law Enforcement (LE), and Road Patrol (RP) applications. The dataset can be accessed at: AOLP Dataset (HyperAI). The dataset does not provide a DOI and is publicly available for academic research purposes. The dataset was accessed on 20 March 2026. The implementation code for the proposed = model has been publicly archived on Zenodo and is available at: https://doi.org/10.5281/zenodo.20640840

Conflict of interest: The authors declare that they have no conflict of interest.

Funding : No funding

Reference

1. W. Khallouli, M. S. Uddin, A. Sousa-Poza, J. Li, and S. Kovacic, “Leveraging transformer-based OCR model with generative data augmentation for engineering document recognition,” *Electronics*, vol. 14, no. 1, Art. no. 5, Jan. 2025, doi: 10.3390/electronics14010005.

2. X. Luan, J. Zhang, M. Xu, W. Silamu, and Y. Li, “Lightweight scene text recognition based on transformer,” Sensors, vol. 23, no. 9, Art. no. 4490, May 2023, doi: 10.3390/s23094490.
3. A. Nagasubramanian, A. S. Almazyad, S. A. Balakrishnan, B. R. M. M, D. Potti, M. Zakariah, and D. M. AlSekait, “OCRNet: A robust deep learning framework for alphanumeric character recognition to assist the visually impaired,” Scientific Reports, vol. 15, Art. no. 41344, Oct. 2025, doi: 10.1038/s41598-025-25278-9.
4. S. Hussain, “Advancing optical character recognition (OCR) with transformer-based architectures,” Scientific Reports, vol. 15, no. 1, pp. 1–15, 2025, doi: 10.1038/s41598-025-25278-9.
5. R. Buoy, M. Iwamura, S. Srun, and K. Kise, “ViTSTR-Transducer: Cross-attention-free vision transformer transducer for scene text recognition,” J. Imaging, vol. 9, no. 12, Art. no. 276, Dec. 2023, doi: 10.3390/jimaging9120276.
6. A. Afkari-Fahandari, E. Shabaninia, F. Asadi-Zeydabadi, and H. Nezamabadi-Pour, “A comprehensive survey of transformers in text recognition: Techniques, challenges, and future directions,” ACM Computing Surveys, vol. 58, no. 5, Art. no. 111, Nov. 2025, doi: 10.1145/3771273.
7. Y. Li, D. Chen, T. Tang, and X. Shen, “HTR-VT: Handwritten text recognition with vision transformer,” Pattern Recognition, vol. 158, Art. no. 110967, 2025, doi: 10.1016/j.patcog.2024.110967.
8. X. Yang, W. Silamu, M. Xu, and Y. Li, “Display-semantic transformer for scene text recognition,” Sensors, vol. 23, no. 19, Art. no. 8159, Sep. 2023, doi: 10.3390/s23198159.
9. Y. Al-Arbo, H. F. Mahmood, and A. Alqassab, “End-to-end license plate detection and recognition in Iraq using a detection transformer and OCR,” Mesopotamian Journal of Computer Science, vol. 2025, pp. 329–340, Sep. 2025, doi: 10.58496/MJCSC/2025/21.
10. L. Wang, C. Cao, B. Zou, J. Ye, and J. Zhang, “License plate recognition via attention mechanism,” *Computers, Materials & Continua*, vol. 75, no. 1, pp. 1801–1814, 2023, doi: 10.32604/cmc.2023.032785.
11. M. Li, Z.-Q. Wang, Y.-H. Zhao, and Q. Li, “License plate recognition for smart construction sites based on GMH-YOLO,” *IEEE Access*, vol. 13, pp. 89760–89774, 2025, doi: 10.1109/ACCESS.2025.3569101.
12. A. Sarhan, R. Abdel-Rahem, B. Darwish, A. Abou-Attia, A. Sneed, S. Hatem, A. Badran, and M. Ramadan, “Egyptian car plate recognition based on YOLOv8, Easy-OCR, and CNN,” *Journal of Electrical Systems and Information Technology*, vol. 11, Art. no. 32, Aug. 2024, doi: 10.1186/s43067-024-00156-y.
13. B. Satya, D. Manongga, H. Hendry, and A. Aminuddin, “Optimized YOLOv8 for automatic license plate recognition on resource constrained devices,” *Engineering, Technology & Applied Science Research*, vol. 15, no. 2, pp. 21976–21981, 2025, doi: 10.48084/etasr.9983.
14. G. R. Gonçalves, S. P. G. da Silva, D. Menotti, and W. R. Schwartz, “A benchmark for license plate character segmentation,” Engineering, Technology & Applied Science Research, vol. 15, no. 2, pp. 21976–21981, 2016.
15. Y. Mu, M. Maimaiti, M. Xu, W. Li, and W. Silamu, “Enhancing scene text recognition with encoder–decoder interactive model,” *Sensors*, vol. 25, no. 24, Art. no. 7684, Dec. 2025, doi: 10.3390/s25247684.
16. N. Plavac, S. A. Amirshahi, M. Pedersen, and S. Triantaphillidou, “Performance of automatic license plate recognition systems on distorted images,” *Journal of Imaging Science and Technology*, vol. 68, no. 6, Art. no. 060401, Nov.–Dec. 2024, doi: 10.2352/J.ImagingSci.Technol.2024.68.6.060401.
17. M. Wu *et al*., "Enhanced PPE detection in low-light tunnel environments: a YOLOv5-based approach," The Visual Computer, vol. 42, no. 1, pp. 8, 2025. DOI: 10.1007/s00371-025-04233-9
18. C. Hao *et al*., "Enhancing scene understanding via attention-guided pyramid fusion recognition under complex imagery conditions," The Visual Computer, vol. 42, no. 1, pp. 36, 2026.
19. H. Shah *et al*., "Advanced driver assistance system (ADAS) and machine learning (ML): The dynamic duo revolutionizing the automotive industry," Virtual Reality & Intelligent Hardware, vol. 7, no. 3, pp. 203–236, 2025.
20. X. Zhang *et al*., "EAPT: Efficient Attention Pyramid Transformer for Image Processing," IEEE Transactions on Multimedia, vol. 25, pp. 50–61, 2023.
21. Rui Liu, Fangbo Lu, Wanchuang Luo, Tianjian Cao, Hailian Xue, Meili Wang, YOLOv8-HAC: Safety Helmet Detection Model for Complex Underground Coal Mine Scene," Computer Animation and Virtual Worlds, vol. 36, no. 4, e70051, 2025.