Fog-Resilient Bangla Car Plate Recognition using Dark Channel Prior and YOLO

Hamim Ibne Nasim[1], Fateha Jannat Printia[1], Mahamudul Hasan Himel[1],
Rubaba Rashid[1], Iffat Jahan Chowdhury[1], Joyanta Jyoti Mondal[2],
Md. Farhadul Islam[1], Jannatun Noor[1]
[1]Computing for Sustainability and Social Good (C2SG) Research Group,
BRAC University, Dhaka, Bangladesh
[2]University of Alabama at Birmingham, Birmingham, AL, USA
{hamim.ibna.nasim, fateha.jannat.printia1, mahamudul.hasan.himel,
rubaba.rashid, iffat.jahan.chowdhury}@g.bracu.ac.bd, jmondal@uab.edu,
{farhadul.islam, jannatun.noor}@bracu.ac.bd

Abstract

Despite advancements in Automatic License Plate Detection (ALPD) methods, the majority of them fail to address the diverse image challenges faced in real-world driving scenarios. These challenges encompass issues like low image quality, contrast issues, etc. Factors such as license plate background, horizontal tilt, and adverse weather conditions like rain or fog further impede LP detection and recognition. This research focuses on the localization and recognition of Bangla vehicle plates in foggy conditions through the application of the Dark Channel Prior (DCP) fog-dehazing technique. The selection of Bangla as the target language is motivated by its status as a low-resource language with high digital text complexity, resulting in limited available resources. The proposed method comprises three main phases. The DCP dehazing algorithm reduces fog in input images initially. Then, the YOLOv8 object detection model is used to identify Bangla license plates from dehazed images, followed by OCR for text recognition. This study leverages DCP, YOLOv8, and OCR technologies to enhance the identification of Bangla vehicle plates under hazardous conditions, thereby contributing to the improvement of transportation safety, law enforcement, traffic management, and taxation processes.

1. Introduction

Bengali/Bangla, spoken by 228 million as a primary language and 37 million as a second language, ranks fifth among native languages and sixth overall. Despite its significant usage, it is considered a low-resource language [31]. The low-resource nature of the language is attributed to insufficient data, limited efforts in the development of Bengali digital text, and challenges stemming from language complexity. According to Joshi et al. [14], the Bangla language is classified as a class 3 language, where the Mean Reciprocal Rank (MRR) is 0.42899. According to a claim by the Bangladesh Road Transport Authority (BRTA), there are 5.7 million registered cars in Bangladesh [5]. However, the parking lot is disorganized, making it challenging to monitor vehicle entry and exit. Urban parking is difficult due to the high volume of automobiles. Detecting illegal vehicles in Bangladesh is problematic due to issues like CCTV and camera noise, misalignment, and the use of cheap sensors and color formats, as observed in our dataset [28]. Managing LP layouts and illumination is difficult. Additional effort is required to enhance the process for reliability. Transportation, environmental protection, public safety, law enforcement, traffic control, revenue collection, and tolls need car tracking. Rain and fog may make automobile monitoring more difficult. The BRTA mandates Bangla LPs, as seen in Figure 1.

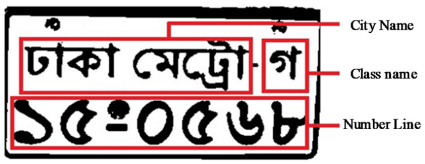

Figure 1. Representation of Vehicle Number Plate in Bangladesh [19]

In our research, we utilize the DCP method to mitigate

haze and fog effects in still pictures. Opting for still images over videos or continuous images to avoid computational challenges and to attain efficiency. We keep videos and continuous images at the forefront of our future work. DCP estimates a depth map, referred to as the dark channel, which is derived from approximately 0.2% of the atmospheric light to characterize light scattering due to atmospheric particles, providing insights into haze levels. Additionally, a transmission map is computed, where pixel values within this map quantify the relationship between the observed image and scene properties, indicating less fog-affected regions. Further refinement is done by the Matting Laplacian matrix of the transmission map, which has been performed through filtering techniques to enhance its quality. Moreover, dehazing improves object detection accuracy, emphasizing the need for clear and high-dynamic-range images. To facilitate the identification of LPs under adverse weather conditions, we employ a custom dataset that underwent rain and fog dehazing procedures.

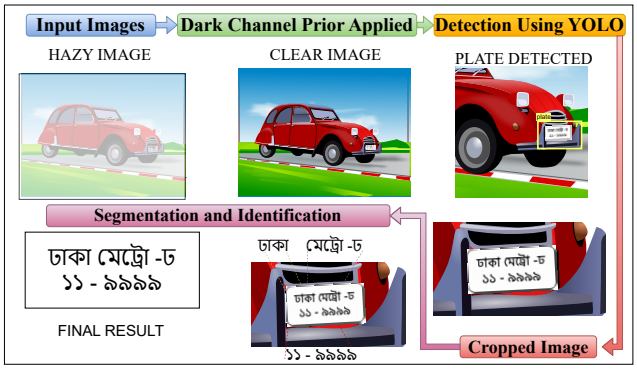

Figure 2. Foggy Car Plate Recognition Workflow

Annotation of this dataset is accomplished using the Computer Vision Annotation Tool (CVAT), and subsequently, the dataset is converted into the YOLO format, encompassing bounding boxes with five associated parameters. Training of our model is conducted using the Ultralytics library, employing the nano version of YOLOv8. This YOLOv8 model is employed for the detection of Bengali car plates, and its performance was assessed across all vehicles within the images. Our results validate the effectiveness of YOLOv8 in accurately detecting objects within images, showcasing its superior performance in the context of vehicle localization. After bounding box localization, image cropping is carried out based on the coordinates obtained from the bounding boxes. The actual characters on the LP have been deciphered with the help of EasyOCR. The cropped images are subsequently processed using EasyOCR to extract and identify characters located within the bounding box regions. Our entire method of implementation is depicted in figure 2.

The following are some of the main contributions of this study:

- We employ depth and density regulation to simulate hazy conditions in standard images.
- We introduce a DCP algorithm designed for dehazing foggy images containing Bengali LPs.

2. Related Works

Within the realm of image and video analysis, the process of License Plate (LP) recognition primarily involves pinpointing and isolating the LP area, which is generally a small fraction of the total visual data, as depicted in Figure 2. Numerous investigations [2, 18, 30] have been dedicated to identifying such compact regions in images in different research domains. Different kinds of operations, such as automated test driving and parking systems, have been met in the transportation sector [17, 22, 25]. Moreover, extensive research has been undertaken to accurately locate a presumed LP zone in photographic or video content.

Besides, Abdullah et al. [1] explore Bangla License Plate Recognition (LPR) in Dhaka using YOLOv3, which segments the input image into a grid to predict bounding boxes and object classes [6]. They adapted YOLOv3 for Bangla LP detection, utilizing pre-trained weights from Darknet-53 layers. The technique efficiently identifies Bangla plates, particularly those with a unique format featuring a character and six numbers. However, the single-scale feature map of YOLOv3 can struggle with varying object sizes and complex scenarios, potentially causing incorrect detections [33].

Moreover, Islam et al. [11] uses a changing boundary to get rid of the horizontal and vertical histogram values. The Region of Interest (ROI) has been taken from a picture of a license plate. This method may be used to locate both textual and non-textual objects. As a result, non-textual components are filtered out using an area and aspect-based strategy. In another study, Azam et al. [4] have applied the Radon transform-based tilt correction approach to rectify tilted LP for the very first time. To get rid of places that are not LP, a new criterion based on picture entropy is being used. The authors of this research suggest a two-type detection pipeline that is integrated to use the Vision API to achieve speedy reasoning in real-time while maintaining consistent accuracy in the detection and recognition of patterns.

Furthermore, Redmon et al. [27] introduce a novel deep learning technique that has been suggested and the single forward computation that allows complete LP identification and recognition. For LP recognition without pre-segmentation, the recognition network uses Gated Recurrent Units (GRU) and Connectionist Temporal Classification (CTC). The CCPD dataset is used to evaluate the suggested technique.

On the other hand, Rahman [26] employ a Convolutional Neural Network (CNN) for Bengali License Plate (LP) recognition, initially cropping the LP from the main image, followed by segmenting digits and characters into 32×32 pixel units. These elements underwent various transformations like scaling and rotation. The study also delved into Dark Channel Prior (DCP) for dehazing unclear images, analyzing its efficacy across four key stages. The research offers detailed step-by-step instructions for DCP-based dehazing, encompassing stages like dark channel generation, ambient light estimation, transmission map creation, and image enhancement [15].

Moreover, Xuesong et al. [13] introduce a DCP method enhancing video quality in low-light and low-visibility conditions. They employ local smoothing and Gaussian pyramid operators with the DCP algorithm to enhance image details. Notably, their approach achieves a higher frame rate of 33 frames per second, distinguishing it from Dong et al.'s [7] method, which produces 23 frames per second, offering a significant improvement for real-time video enhancement.

In addition, Yash et al. [29] suggest using Optical Character Recognition (OCR) and a new method based on deep learning to automatically find and read number plates. Using deep learning, the model is taught to spot the car. The part of the picture that shows the LP is cut out well, and a CNN uses OCR to figure out what the numbers and letters are. The Jetson TX2 NVIDIA target is used to train the model, and a public dataset from the Kaggle database is used to test how well it works.

To the best of our knowledge, no prior research has concentrated on Bangla car number plates and low-resource languages in foggy weather or any form of hazardous weather. So, we move forward on exploring an in-depth study in this area.

3. Dataset

We acquire a secondary dataset consisting of 2,754 samples, which has been partitioned into three subsets: the training set encompasses 70% of the data, comprising 1,928 images, the test set consists of 15%, accounting for 413 images, and the validation set also comprises 15%, containing 413 images. The dataset exhibits a diverse range of characteristics, encompassing variations in image clarity, sharpness, distances between objects, LP attachment status, diurnal and nocturnal scenarios, and other relevant attributes. The dataset comprises a distribution of 46.6% cars, 2.6% LPs alone, 39.4% bikes, 3.3% vehicles utilizing compressed natural gas (CNG), and 8.1% trucks and buses.

3.1. Customized Dataset

We opt to employ a customized dataset, a derivative of a pre-existing secondary dataset initially curated by Sams et al. [28]. This dataset comprises two distinct segments: the first segment comprises images featuring vehicles with their associated LPs, while the second segment exclusively includes images of isolated LPs. Given our specific focus on isolating the LP region and mitigating environmental disturbances, we have selected to utilize the first segment of the dataset, which consists of 1,928 images allocated for training, along with 413 images designated for both validation and testing. Our choice to work with this subset is facilitated by the fact that it is already formatted in the YOLO standard, streamlining its compatibility with our research objectives.

To introduce a simulated foggy effect into the images, we follow the approach described by Godard et al. This method entails the estimation of object depth using a single image and is achieved through self-supervised training, where the model learns to predict the perspective of one image based on another. This process leverages the disparity between the two images as an intermediary variable, allowing for the derivation of a depth map (D_t). Subsequently, an error is quantified according to Equation 1 to ascertain the final image and the associated generation process [8].

$$L_p = \min pe\left(I_t, I_{t' \to t}\right).t' \tag{1}$$

where L_p is the mistake in photometric projection and pre-photometric pe reconstruction error, which is calculated using SSIM and L1 distance.

SSIM and L1 distance are both parts of the photometric restoration mistake, which we get from various research works [9, 34, 36],

$$pe\left(I_a, I_b\right) = \frac{\alpha}{2}\left(1 - \text{SSIM}\left(I_a, I_b\right)\right) + (1-\alpha)\left|\left|I_a - I_b\right|\right| \tag{2}$$

In Equation 2, $\alpha = 0.85$ is the weight of the wall. As in we use edge-aware smoothness. A smoothness term (equation 3) has been added to make sure that the depth maps are smooth.

$$L_s = \left|\partial_x d_t\right| e^{-\left|\partial_x I_t\right|} + \left|\partial_y d_t\right| e^{-\left|\partial_y I_t\right|} \tag{3}$$

According to Godard et al. [8] various techniques, including pre-pixel minimum re-projection, have been employed to enhance the precision of the projected depth maps while refraining from introducing additional models into the framework. To address scenarios involving moving objects, the minimum error and auto-masking stationary strategies were implemented. The binary mask μ in equation 4 is dynamically computed, taking into account projection errors and a multiscale estimation, to prevent convergence to local minima during the optimization process. The collective training loss is obtained by aggregating the per-pixel smoothness term (L_s) with the filtered photometric loss [16, 32, 37].

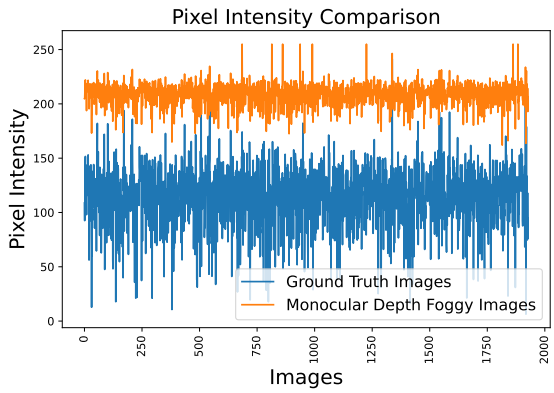

Figure 3. Mean Intensity of Ground Truth Images and Customized Foggy Images

$$\mu = \left[\min_{t'} pe\left(I_t, I_{t'\to t}\right) < \min_{t'} pe\left(I_t, I_{t'}\right)\right] \quad (4)$$

This is how we derived the depth maps from the original images. Subsequently, we inverted these images. Once the model is appropriately configured, we proceed to import the designated target image onto which we intend to simulate the foggy conditions. We also explore methodologies to improve the accuracy of the projected depth maps without the need for additional model complexity. Upon meticulous examination, it becomes evident that the artificially generated foggy images closely resemble their corresponding original counterparts, exhibiting minimal discernible distinctions.

In a conventional haze-free image, the color image comprises pixel values ranging from 0 to 255, serving as indicators of the clarity of each pixel's coloration. An image of moderate clarity tends to exhibit greater prominence of both bright and dark values. Conversely, in cases of foggy imagery, there is an observable prevalence of brighter pixel values and a corresponding decrease in darker values due to the limited visibility caused by the fog. As depicted in Figure 3, it is noteworthy that the mean intensity value of the ground truth image is lower than that of the monocular depth-based foggy image. This discrepancy serves as evidence affirming the presence of authentic fog within the customized dataset.

Following the acquisition of the depth-affected image under foggy conditions, we proceed to amalgamate this depth image with the corresponding clear image. This fusion process is subject to the regulation of fog density levels, governed by a beta regulator. As depicted in Figure 4, the illustration showcases individual images exhibiting varying mean intensity values across distinct beta settings. Notably, at a beta value of 7, the image demonstrates a maximum intensity value that notably exceeds 200, whereas the authentic image typically exhibits mean intensities within

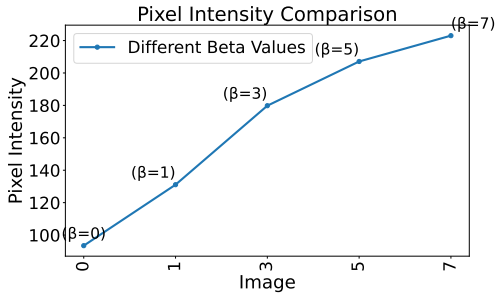

Figure 4. Variation in Density Mean Intensities of the Individual Image

the range of 50 to 100. Since we could not discover any secondary dataset and owing to restrictions in the environment and condition of the current we would like to work with real-world datasets in the future after gathering primary data.

4. Methodology

The methodological framework comprises a tripartite process. Initially, there is a phase dedicated to data preprocessing through the application of the DCP technique. Subsequently, the second stage focuses on the detection of LPs utilizing the YOLO algorithm. Lastly, the third stage encompasses the recognition of the detected LPs through the implementation of EasyOCR, a specialized OCR system. This structured approach is designed to facilitate the comprehensive analysis and identification of LPs in the research context.

4.1. Data Pre-Processing Using DCP

Under hazy atmospheric conditions, this detection process may encounter challenges attributable to particulate matter and fog. These adverse conditions often result in reduced intensity levels of the license plate or a diminished dynamic range, leading to the creation of shadow-like artifacts that adversely affect detection accuracy. Hence, a critical prerequisite for LP recognition in such scenarios involves an initial data preprocessing step.

The data preprocessing operation is fundamentally instrumental in rendering the LP more discernible within the image. In this context, DCP has been employed as a means to address dynamic range limitations and enhance image contrast. The DCP algorithm commences by segregating the input image into its constituent color channels, specifically the green (G), blue (B), and red (R) channels. Subsequently, it calculates the dark channel, which corresponds to the channel featuring the smallest pixel values. To further augment the distinctiveness of the dark channel, a small square kernel is applied through an erosion operation. This preprocessing step serves to accentuate the LP within the

image, rendering it more prominent and thereby facilitating subsequent recognition tasks.

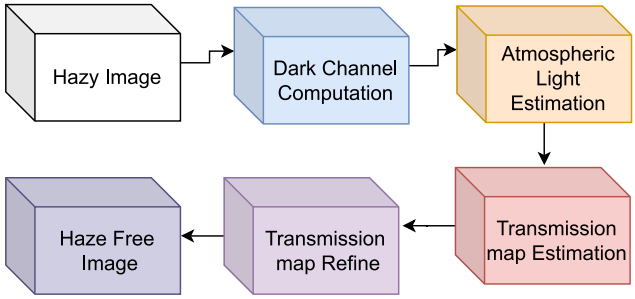

Figure 5. Dark Channel Prior Architecture

4.1.1 Atmospheric Light Analysis

According to He et al. [10], estimating the atmospheric light (A) can be done with a subset of the darkest pixels in the dark channel. It computes the average value of the darkest pixels in the dark channel and the brightest pixels to measure the light, as the atmospheric light has the highest intensities from those pixels. This average serves as the atmosphere's illumination. A_i is specified in equation 5.

$$A_{i=R,G,B}^{n+1} = \frac{1}{2}\left(A_i^n + R\left(x_n + 1\right)\right) \quad x \in \widetilde{R} \tag{5}$$

where $\widetilde{R}$ is the order of the top 0.2% of the brightness value in the dark channel of $R(x)$. The value of A_i is found by comparing and updating the average of the pixel points that are next to each other in $\widetilde{R}$. Because of this, each repetition must be compared to t. Through repetition, the places where the dark channel brightness is not very noticeable are taken into account.

4.1.2 Transmission Map Forecast

The transmission map (t) is calculated using the atmospheric light channel and the dark channel. Here, $\tilde{t}(\mathbf{x})$ in equation 6 is determined by the expression.

$$\tilde{t}(\mathbf{x}) = 1 - \min c\left(\min \mathbf{y} \in \Omega(\mathbf{x})\left(\frac{I^c(\mathbf{y})}{A^c}\right)\right) \tag{6}$$

where I^c is the normalized input image divided by the atmospheric light A^c and omega is a constant. In fact, $\min c\left(\min \mathbf{y} \in \Omega(\mathbf{x})\left(\frac{I^c(\mathbf{y})}{A^c}\right)\right)$ is the dark channel of the normalized haze image $\frac{I^c(\mathbf{y})}{A^c}$. It directly provides an estimation of the transmission.

4.1.3 Transmission Refinement via Guided Filtration

According to He et al. [10], using directed filtering to refine the transmission map enhances edges and maintains essential details. As inputs, the guided filter utilizes the grayscale image and the estimated transmission map to generate a refined transmission map (t). $t(\mathbf{x})$ stands for the refined transmission map. We minimize the following cost function by writing $t(\mathbf{x})$ and $\tilde{t}(\mathbf{x})$ as t and $\tilde{t}$ in the form of a vector.

$$E(\mathbf{t}) = \mathbf{t}^T\ \mathrm{L}\mathbf{t} + \lambda(\mathbf{t} - \tilde{\mathbf{t}})^T(\mathbf{t} - \tilde{\mathbf{t}}). \tag{7}$$

The equation 7 is the cost function, where L is the Matting Laplacian matrix that had been proposed by Levin. Here, λ is a regularization parameter.

4.1.4 Image Restoration

The process involves the reconstruction of the dehazed image through the utilization of the improved transmission map (t), the atmospheric light (A), and a conservative threshold value. Adjustments and shifts in pixel values for each color channel are carried out to regain the dehazed image (J), a technique described by He et al. [10]. Upon examination of the intensity frequency distribution before and after the application of the DCP, it becomes evident that the dynamic range has undergone alteration. Specifically, post-application of the DCP, the intensity values exhibit a more dispersed distribution compared to the initial state, wherein the image displayed greater prominence.

Figure 6. Pre-processing using DCP

4.2. Models

Our experimental setup involve the selection of four distinct models, in conjunction with a proposed model, to investigate their efficacy in the context of image enhancement. The first model under consideration is the Multi-scale Retinex, a method that traditionally utilizes an array of bandpass filters to decompose the image into multiple scales. Subsequently, it employs local contrast adjustment techniques, including logarithmic transformations and divisive normalization, to enhance image quality. Additionally, we leverage the DCP technique, which plays a pivotal role

in estimating the hazy map within the images. However, unlike conventional non-local dehazing approaches, DCP employs a CNN to disentangle the haze and enhance visibility. The third model incorporated into our analysis is Fast Visibility Restoration, a method designed to dehaze single images through a combination of techniques, including white balance correction, atmospheric veil estimation, corner preservation smoothing, and specialized tone mapping.

Our experimental evaluation is based on three key parameters: Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), and Execution Time. SSIM provides a measure of structural similarity between the modified image and the original, encompassing luminance, contrast, and structural components. The model exhibiting the SSIM value closest to the ground truth is deemed the most representative. PSNR, on the other hand, quantifies image quality relative to the ground truth. Lastly, execution time denotes the duration required for processing a dataset using a particular algorithm.

Our findings reveal that among the considered models, DCP exhibits the highest SSIM and PSNR values, measuring 0.81 and 19.71, respectively. Notably, Fast Visibility Restoration demonstrates significantly faster execution, outperforming DCP by a factor of 15.5. However, it is important to note that rapid visibility adjustments in the latter model may result in deviations in hue, saturation, and brightness compared to real-world colors, which constitute a primary challenge.

Dehazing Algorithms	**Average SSIM**	**Average PSNR**	**Execution Time(s)**
DCP	0.8188	19.7172	0.4257
Multi-Scale Retinex	0.7777	14.4045	7.3202
Non-local Image Dehazing	0.7444	15.4989	10.5618
Fast Visibility Restoration	0.7377	16.879	0.027

Table 1. SSIM Score, PSNR Score, and Execution Time Per Picture of Four Dehazing Models

Upon examination of Figure 7, it becomes evident that the graphical representations therein depict the contrast enhancement outcomes achieved by various algorithms relative to the original hazy image. Among the algorithms evaluated, the most substantial contrast enhancement is observed in the results obtained through the utilization of the Multi-scale Retinex and DCP methods. The term "contrast enhancement" pertains to the disparity between the minimum and maximum pixel values within the image. However, when assessing the accuracy of haze removal using the Structural Similarity Index (SSIM), it is noteworthy that

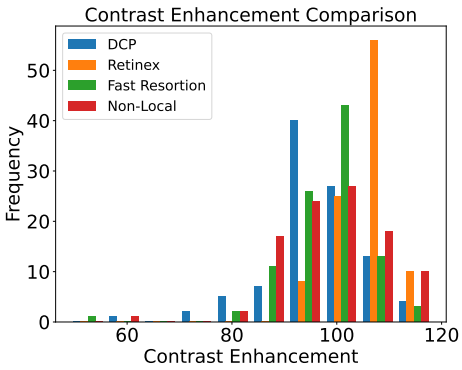

Figure 7. Comparison: Contrast Enhancement and Haze Removal Accuracy

the DCP method attains the highest accuracy level, measuring at 80%. This achievement is further underscored by the frequency at which such superior accuracy is consistently observed in comparison to alternative methodologies.

Figure 8. Visual Comparison of Single Image Dehazing Algorithms

Figure 8 shows the individualized foggy representation alongside the ground truth. and compare the output of the four individual image dehazing algorithms on the same image. The visualization shows us which luminance, hue, and saturation values are most significantly shown on DCP compared to others.

4.3. Detection Using YOLOv8

We utilize a custom dataset comprising over 2,754 images, specifically focusing on hazy weather conditions. Dehazing of the images is executed using the DCP algorithm, followed by meticulous data annotation involving the labeling of each LP within every image. This dataset is subsequently transformed into YOLO 1.1 format, with annotated images stored as text files, each bounding box represented by five parameters signifying class, center, width, and height.

The model is trained using the Ultralytics library, specif-

ically the nano version of YOLOv8. The training configurations, encompassing class names, file paths, training settings, and validation datasets, are defined within a YAML file. The model effectively detects and localizes all vehicles within the images, denoted by the red bounding boxes. Subsequently, the model undergoes rigorous testing with new data, unequivocally showcasing the superior performance of the YOLOv8 object identification model in detecting and localizing objects within images.

In the experimental phase, five distinct models are employed to identify LPs. The initial iteration of YOLO involved bounding box detection to discern licensed content, with multiple YOLO variants implemented on separate backbones. These YOLO iterations encompass v8n, v8m, v5n, and v5m. Notably, in terms of parameters and floating-point operations per second, YOLOv8m exhibit twice the computational potency of YOLOv8n. For YOLOv8m, we use CSPdarknet50 as the backbone and 100 epochs to train our dataset for detection. For other models, optimal parameters have been used, which may outperform.

A comprehensive evaluation is conducted on 15% of the dataset images following a predefined training and validation period. Results indicate that YOLOv8m achieves the highest accuracy at 0.985, a commendable F1 score of 0.96, a recall confidence of 0.98, and precision at 1.00. Specifically, YOLOv8m demonstrates an accuracy of 0.985 and a recall rate of 0.98 at a confidence threshold of 0.430, surpassing the performance of YOLOv5n and YOLOv5m, which achieved 0.971 and 0.982, respectively, and YOLOv8n with a score of 0.942. The hyperparameters for YOLO8m are - patience = 50, batch size = 16, learning rate initial = 0.01, and learning rate final = 0.01. Since we have an adequate amount of samples and our main contribution is in the form of data pre-processing, we skip the use of augmentation.

Detection Models	Accuracy	F1 Score	Precision Score	Recall Score
YOLOv8m	**0.985**	**0.96**	1.00	0.98
YOLOv5n	0.971	0.96	1.00	0.98
YOLOv8n	0.942	0.94	1.00	0.98
YOLOv5m	0.982	0.96	1.00	**0.99**

Table 2. Performance Metrics of Detection Models

5. Recognition of Character Using EasyOCR

The text outlines the procedure for configuring an OCR reader for Bengali text recognition via EasyOCR. It incorporates object detection through the deployment of YOLOv8, which identifies objects within designated images. When an object is detected, the corresponding Region of Interest (ROI), typically an LP, is extracted. Subsequently, OCR analysis is executed on these cropped images, primarily containing LPs. EasyOCR supports over 70 languages and many fonts and text styles. Fast and real-time, EasyOCR processes massive photo volumes. Its interface is easy to integrate into Python apps [12]. EasyOCR excels at OCR benchmarks. The OCR outcomes are managed through the extraction of recognized text from the cropped images. The process includes tracking total confidence scores and character counts for the subsequent calculation of an average confidence score, computed by dividing the cumulative confidence score by the count of recognized characters.

Models	Average confidence
YOLOv5m	0.6648
YOLOv8n	0.6551
YOLOv5n	0.6497
YOLOv8m	0.6843

Table 3. EasyOCR Average Confidence Score in Different Models

Figure 9. OCR Recognition: True vs. False

EasyOCR is able to identify characters in 2237 out of a total of 2754 photos (81.23%). EasyOCR demonstrates a favorable ability to accurately detect most individual characters; however, when considered collectively, its performance exhibits a range from lower to moderate proficiency. As illustrated in Figure 9, instances of both true positives and false positives in EasyOCR recognition are evident. Notably, an observed advantage is that an increase in image enhancement through the application of DCP corresponds to improved OCR capabilities.

Limitation. Limitations arise when dealing with low-resolution images, contributing to reduced accuracy. These limitations may stem from factors such as 1) inadequate image resolution; 2) attachment of obstructions to the number plate; 3) character discoloration; 4) color matching between the plate and characters; and 5) the presence of the BRTA seal.

6. Discussion

To mitigate haze in the images, several algorithms, including Multi-Scale Retinex, Non-Local Image Dehazing,

Fast Visibility Restoration, and DCP, are employed. Notably, the DCP method outperforms other models in terms of SSIM and PSNR, indicating a higher degree of structural and color similarity between the enhanced images and the ground truth images, compared to alternative approaches.

Several factors contribute to the superior performance of DCP. Firstly, the dark channel effectively encapsulates the depth of haze within the image, in contrast to Multi-Scale Retinex, which focuses on luminance and color. The conversion of illumination components from Retinex into grayscale images diminishes its PSNR and SSIM values. In the case of Fast Visibility Restoration, color alterations result from the conversion from RGB to LAB color space. Conversely, in Non-Local Image Dehazing, reliance on the transmission map impacts its performance. DCP, on the other hand, employs Laplace Matting for image restoration, while Non-Local Image Dehazing involves refinement, which results in a loss of structural information and saturation [35].

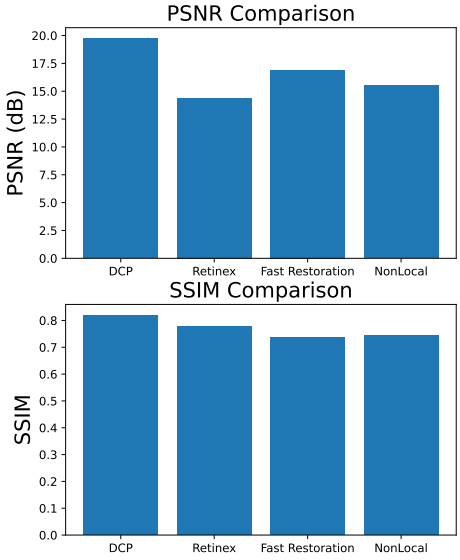

Figure 10. SSIM and PSNR Score of All Models

The superior performance of YOLOv8m over YOLOv8n, YOLOv5m, and YOLOv5n, particularly with the CSPDarknet50 backbone, can be attributed to several factors. Notable enhancements, including the introduction of anchor box creation, a refined loss function, and the incorporation of the Spatial Pyramid Pooling (SPP) module, are distinctive features introduced in YOLOv8. These advancements likely conferred advantages to YOLOv8m, contributing to its superior performance.

Additionally, YOLOv8m benefits from a more extensive and diversified training dataset compared to YOLOv8n, YOLOv5m, and YOLOv5n. This broader dataset facilitates a deeper understanding of the surrounding environment and the ability to generalize from new information. Furthermore, the superior hyper-parameter settings of YOLOv8m relative to v8n, v5m, and v5n likely play a pivotal role in its heightened performance. It is also noteworthy that the adoption of the novel CSPDarknet50 backbone architecture in YOLOv8m potentially confers a performance advantage over its counterparts, YOLOv8n, YOLOv5m, and YOLOv5n.

7. Conclusion and Future Work

In the research, we present a novel approach to Automatic License Plate Detection (ALPD) tailored for challenging foggy conditions, utilizing the Dark Channel Prior (DCP) dehazing technique combined with the advanced YOLOv8 object detection model. This methodology is specifically designed to improve the detection and recognition of Bangla car plates, a language often overlooked in digital text complexity due to its low-resource status. The DCP algorithm significantly enhances image clarity by effectively reducing fog, a critical step in ensuring the visibility and accuracy of license plate detection. YOLOv8 provides a highly effective means of identifying Bangla license plates from the dehazed images. This approach has proven exceptionally effective, achieving an impressive license plate identification accuracy of 98.5%.

Following the detection process, Optical Character Recognition (OCR) technology is employed to decipher and extract text from the license plates, achieving a confidence score of 68.43%. Incorporating this kind of technologies into multimedia-cloud storage [20, 21, 23, 24], especially within a resource-constrained environment not only advances the domains of transportation safety and law enforcement in the developing countries but also holds significant importance in bolstering the safety of vehicle passengers. This concern is particularly noticeable in Bangladeshi cities, as reported by a recent study [3].

Looking forward, the research aims to extend its focus to Hindi, another language with limited digital resources. Our future endeavors will concentrate on refining image quality through techniques like contrast-limited Adaptive Histogram Equalization (CLAHE) and exploring high-contrast black-and-white imaging. Additionally, we aim to improve detection capabilities using models like DETR and to resolve challenges associated with OCR, particularly in reading handwritten text and deciphering serial scrambling. A particular emphasis will be placed on optimizing ALPD for low-light and nighttime conditions, addressing both image quality and object detection challenges, thereby broadening the applicability and reliability of our proposed method in diverse and challenging driving environments.

References

[1] S. Abdullah, M. Mahedi Hasan, and S. Muhammad Saiful Islam. Yolo-based three-stage network for bangla license plate recognition in dhaka metropolitan city. In *2018 International Conference on Bangla Speech and Language Processing (ICBSLP)*, pages 1–6, 2018. 2

[2] Tanzia Ahmed, Tanvir Rahman, Bir Ballav Roy, and Jia Uddin. Drone detection by neural network using GLCM and SURF features. *J. Inf. Syst. Telecommun.*, 9(33):15–24, 2021. 2

[3] Aditto Baidya Alok, Hasibul Sakib, Shamsil Arafin Ullah, Fardin Huq, Riya Ghosh, Joyanta Jyoti Mondal, Md. Sadiqul Islam Sakif, and Jannatun Noor. 'khep': Exploring factors that influence the preference of contractual rides to ride-sharing apps in bangladesh. In *Proceedings of the 6th ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies*, COMPASS '23, page 43–53, New York, NY, USA, 2023. Association for Computing Machinery. 8

[4] S. Azam and M. M. Islam. Automatic license plate detection in hazardous condition. *J. Vis. Commun. Image Represent.*, 36:172–186, 2016. 2

[5] Bangladesh Road Transport Authority (BRTA). Number of registered vehicles in bangladesh, 2023. 1

[6] Datacamp.com. Yolo object detection explained, 2022. 2

[7] X. Dong et al. Fast efficient algorithm for enhancement of low lighting video. In *2011 IEEE International Conference on Multimedia and Expo*, pages 1–6, 2011. 3

[8] Clément Godard, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. Digging into self-supervised monocular depth estimation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 3828–3838, 2019. 3

[9] Clément Godard, Oisin Mac, A. Gabriel, and J. Brostow. Unsupervised monocular depth estimation with left-right consistency. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 3

[10] K. He, J. Sun, and X. Tang. Single image haze removal using dark channel prior, 2009. 5

[11] R. Islam, M. R. Islam, and K. H. Talukder. An efficient method for extraction and recognition of bangla characters from vehicle license plates. *Multimed. Tools Appl.*, 79(27-28):20107–20132, 2020. 2

[12] A. Jain. Comparing ocr methods: Paddle ocr vs tesseract ocr vs easyocr vs kerasocr. *Medium*, 2023. 7

[13] Xuesong Jiang, Hongxun Yao, Shaojie Zhang, Xiaoyan Lu, and Wenming Zeng. Night video enhancement using improved dark channel prior. In *2013 IEEE International Conference on Image Processing*, pages 553–557, 2013. 3

[14] P. Joshi, S. Santy, A. Budhiraja, K. Bali, and M. Choudhury. The state and fate of linguistic diversity and inclusion in the nlp world, 2020. 1

[15] S. Lee, S. Yun, J.-H. Nam, C. S. Won, and S.-W. Jung. A review on dark channel prior based image dehazing algorithms. *EURASIP Journal on Image and Video Processing*, 2016(1), 2016. 3

[16] Cheng Luo and et al. Every pixel counts ++: Joint learning of geometry and motion with 3d holistic understanding. *arXiv preprint arXiv:1808.03867 [cs.CV]*, 2018. 3

[17] Md Imtiaj Miah, Joy Chandra Gope, Ananta Deb Nath, AKM Janaite Nain, Faria Nasir Mitu, and Jannatun Noor. Advanced waterway transport system based on internet of things (iot): A novel approach. In *2022 25th International Conference on Computer and Information Technology (ICCIT)*, pages 1–6. IEEE, 2022. 2

[18] Joyanta Jyoti Mondal, Md. Farhadul Islam, Sarah Zabeen, and Meem Arafat Manab. Invopotnet: Detecting pothole from images through leveraging lightweight involutional neural network. In *2022 25th International Conference on Computer and Information Technology (ICCIT)*, pages 599–604, 2022. 2

[19] N.-A.-Alam, M. Ahsan, M. A. Based, and J. Haider. Intelligent system for vehicle number plate detection and recognition using convolutional neural networks. *Technologies (Basel)*, 9(1):9, 2021. 1

[20] Jannatun Noor, Hasan Ibna Akbar, Ruhul Amin Sujon, and ABM Alim Al Islam. Secure processing-aware media storage (spms). In *2017 IEEE 36th International Performance Computing and Communications Conference (IPCCC)*, pages 1–8. IEEE, 2017. 8

[21] Jannatun Noor and ABM Alim Al Islam. ibuck: Reliable and secured image processing middleware for openstack swift. In *2017 International Conference on Networking, Systems and Security (NSysS)*, pages 144–149. IEEE, 2017. 8

[22] Jannatun Noor, Md Golam Hossain, Muhammad Ahad Alam, Ashraf Uddin, Sriram Chellappan, and ABM Alim Al Islam. Svload: An automated test-driven architecture for load testing in cloud systems. In *2018 IEEE Global Communications Conference (GLOBECOM)*, pages 1–7. IEEE, 2018. 2

[23] Jannatun Noor, Saiful Islam Salim, and ABM Alim Al Islam. Strategizing secured image storing and efficient image retrieval through a new cloud framework. *Journal of Network and Computer Applications*, 192:103167, 2021. 8

[24] Jannatun Noor, Md. Nazrul Huda Shanto, Joyanta Jyoti Mondal, Md. Golam Hossain, Sriram Chellappan, and A. B. M. Alim Al Islam. Orchestrating image retrieval and storage over a cloud system. *IEEE Transactions on Cloud Computing*, 11(2):1794–1806, 2023. 8

[25] Salma Hani Nova, Shafquat Mahmud Quader, Saumitra Das Talukdar, Mubtasimur Rahman Sadab, Md Shaba Sayeed, ABM Alim Al Islam, and Jannatun Noor. Iot based parking system: Prospects, challenges, and beyond. In *2022 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)*, pages 393–400. IEEE, 2022. 2

[26] M. M. Shaifur Rahman, M. Mostakim, M. S. Nasrin, and M. Z. Alom. Bangla license plate recognition using convolutional neural networks (cnn). In *2019 22nd International Conference on Computer and Information Technology (ICCIT)*, pages 1–6, 2019. 3

[27] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. You only look once: Unified, real-time object detection. *arXiv [cs.CV]*, 2015. 2

[28] Ataher Sams and Homaira Huda Shomee. Bangla LPDB - a, 2021. 1, 3

[29] Y. Shambharkar, S. Salagrama, K. Sharma, O. Mishra, and D. Parashar. An automatic framework for number plate detection using ocr and deep learning approach. *International Journal of Advanced Computer Science and Applications*, 14(4), 2023. 3

[30] Labib Ahmed Siddique, Rabita Junhai, Tanzim Reza, Salman Sayeed Khan, and Tanvir Rahman. Analysis of real-time hostile activitiy detection from spatiotemporal features using time distributed deep cnns, rnns and attention-based mechanisms. In *2022 IEEE 5th International Conference on Image Processing Applications and Systems (IPAS)*, volume Five, pages 1–6, 2022. 2

[31] The Business Standard. Languages of bangladesh, 2022. 1

[32] Sudheendra Vijayanarasimhan, Susanna Ricco, Cordelia Schmid, Rahul Sukthankar, and Katerina Fragkiadaki. Sfm-net: Learning of structure and motion from video. *arXiv preprint arXiv:1704.07804 [cs.CV]*, 2017. 3

[33] X. Wang, J. Liu, and X. Zhu. Early real-time detection algorithm of tomato diseases and pests in the natural environment. *Plant Methods*, 17(1):43, 2021. 2

[34] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 3

[35] L. Wu et al. Hybrid dark channel prior for image dehazing based on transmittance estimation by variant genetic algorithm. *Applied Sciences (Basel)*, 13(8):4825, 2023. 8

[36] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017. 3

[37] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6612–6619, 2017. 3