This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3401044

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

A Robust Intelligent System for Text-Based Traffic Signs Detection and Recognition in Challenging Weather Conditions

Sara Khalid[1], Jamal H. Shah[1], Muhammad Sharif[1], Fadl Dahan[2], Rabia Saleem[3], Anum Masood[4*]

[1] Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan.

[2] Department of Management Information Systems, College of Business Administration Hawtat Bani Tamim, Prince Sattam Bin Abdulaziz University, Al-Kharj 11942, Saudi Arabia.

[3] Department of Information Technology, Government College University Faisalabad, Pakistan.

[4] Department of Circulation and Medical Imaging, Norwegian University of Science and Technology, Trondheim, Norway.

[*] Corresponding (anum.masood@ntnu.no)

ABSTRACT Traffic signs have great importance regarding smooth traffic flow and safe driving. However, due to many distractions and capricious factors, spotting and perceiving them may become hazardous. Traffic sign detection and recognition have gained popularity to put an end or to lessen the issue, and massive efforts have been realized in this regard. Despite considerable endeavors put together for traffic sign detection and recognition, there is a lack of attention in this area where these traffic signs contain text in them. A handful of studies may be found in state-of-the-art (SOTA) methods for text-based traffic sign detection, and particularly lesser for text recognition of detected text. The proposed method focuses on developing a robust semi-pipeline intelligent system to detect and understand text from traffic road signs boards in various weather conditions. For this purpose, a customized YOLOv5s is deployed for initial panel detection. Subsequently, MSER with preprocessing techniques is used for localization of text. Finally, OCR with NLP is utilized to recognize the text. The proposed method employed the ASAYAR dataset for training and different datasets for testing. The proposed approach produced satisfactory outcomes on them in contrast with SOTA approaches.

INDEX TERMS Text Recognition; Deep Learning; Natural Language Processing; Automated Road Signs/Panels Detection; YOLOV5s; MSER.

I. INTRODUCTION

Every country has traffic signs for better traffic flow and for the safety of drivers and pedestrians. Therefore, signs play crucial role every civilized country. Traffic signs encompass of symbols, text or both with particular shapes and distinguishing colors. Text-based traffic signs provide semantic context based on environment, offering higher accuracy in contrast to traffic signs based solely on symbols. The traffic signs containing text may have directional or other information about road which a high-definition camera mounted on a car may capture. The directional panel or sign board on the road contains text, arrows, numbers or combination of all to indicate a route or a turn. The primary objection of identifying traffic signs having symbols only is paramount but text-based signs are more significant, as they aid drivers in avoiding grave situation, for automobile navigation and providing hope for visually impaired people. Detection and recognition of traffic signs both text-based and symbol based have significant value in systems related to vehicles and driving by playing crucial role in both auxiliary and autonomous driving, like traffic congestion detection [1], road detection [2, 3], etc. Similarly, detection of text-based traffic panel or sign may help map companies or traffic authorities to automatically and rapidly gathering roads and relevant information at a lower amount of cost.

Nevertheless, detecting traffic signs containing symbol is frequently the foremost concern in the research field [4-9] such as circular signs detection, speed limit sign detection, stop sign detection and so on. Few studies have been conducted on the detection of traffic signs with text or text-based traffic panels [10-15] and one of the major challenge among others is the diversity of text in terms of font, scale, rotation, orientation, blur, dirt, haze and illumination which affects accuracy and time cost [13, 16].

Furthermore, differences in language scripts among regions have contributed greatly in scarcity of algorithms for detecting text based road panels and traffic signs [17]. The inclusion of text within the backgrounds of traffic signs, as well as text-based traffic panels, is another obstacle in detection and based on this, the traffic signs images can be categorized as wild and express road images and street-level images [18]. The textual traffic sign images from wild (less complex open areas), highway and express categories have slightly simpler backgrounds like large area of sky, huge mountains, myriad trees, making them somewhat more manageable. Predominantly, research has been caried out on images from wild, highway and express categories with the simpler backgrounds. However, traffic signs with such backgrounds are still being researched due to scale variability, similar object presence and occlusion issues [19].

The textual traffic signs with street-level background imagery often feature crowded pedestrian areas, buildings, other vehicles and various similar objects such as billboards. These factors may lead to false detection when alike objects are misapprehended for text-based traffic signs. The detection in such images is difficult and less work can be seen in this area.

(a) (b) (c) (d) (e) (f) (g) (h)

FIGURE 1. Text-based Traffic Signs with Varying Resolution, Conditions, and Settings: (a) in Wild; (b) and (c) at Street Level; (d) and (e) at Night; (f) in Cloudy Weather; (g) in Fog; (h) in Smog

Additionally, text detection itself possess a substantial challenge and a popular topic in vision systems. The detection of text in road signs falls under the category of detecting scene text, which is a subcategory of detecting text in scene images [20, 21]. Text detection in outdoor images faces many challenges, including inclement climate change, such as obscure hazy environment which cause image degradation and deterioration. Additionally, aerosols such as mist, dust, smoke, etc. producing noise in images. Furthermore, atmospheric amalgamation and dispersion trigger addition of air light attenuation while camera is capturing scene and resulting in washed out image [22]. These issues may result in a poor quality and low contrast images with faded, partially occluded, and disoriented objects, from which detecting and recognizing objects become more challenging. Text detection, in particular, turns out to be even more difficult under these conditions.

The properties and attributes of the background and foreground of an image put forward another major challenge in text extraction and recognition from an image [23]. Furthermore, text detection in traffic signs becomes more difficult in night times due to darkness and challenges related to light geometry as presented in image in FIGURE 1. All these obstacles make the task of detecting text in traffic signs and text-based traffic panels even more exigent. Recently, many formerly unsolved and unheeded issues like detection of text in text-based traffic signs, have become gradually easier to address with the advent of new technologies such as deep learning. Text detection in the wild (open and less complex background scenes) is quite successful by using deep learning methods. Nevertheless, due to discrepancy in text, deep learning methods require a gigantic amount of data to achieve better outcomes, causing for considerable time consumption. Besides, time costly methods are not suitable for real time environment-based systems. Consequently, we have proposed an efficient and effective algorithm using deep learning to detect text-based traffic guided panels and to recognize the text in diverse situations with MSER and OCR with NLP.

II. OBJECTIVE AND CONTRIBUTION

The objectives and contributions of proposed method are summarized as:

- Time efficient algorithm meeting real time requirement of ITS.
- A robust algorithm to detect and recognize text in road signs and panels in different situations
- Improving the accuracy with reducing false positive rate

III. LITERATURE REVIEW

A brief literature review is presented in two sections. The first section discusses traffic sign detection and recognition of symbol-based traffic signs. The second section not only gives a short insight on traffic sign detection and recognition of text-based traffic signs but also offers insight into scene text detection.

A. TRAFFIC SIGN DETECTION AND RECOGNITION (TSDR)

For decades, researchers have been working and focusing on developing a stable, generic and time efficient TSDR to mitigate many cumbersome situations and to achieve autonomous ITS. Primarily, TSDR are divided in two [24] or sometimes three phases [4] of detection, tracking and recognition. The detection phase concentrates on locating traffic signs, the tracking phase handles the information regarding video sequences, and the recognition phase comprehends the content of traffic sign.

The traffic signs may contain symbol, text or both. Therefore, a detection method should be capable to handle both symbol and text. However, in actual, both problems are addressed independently by researchers. Traffic signs containing symbol have been researched extensively [4-8, 25], while text based traffic signs have received relatively less attention, regardless of the information they deliver [10-14]. The factors that have made symbol-based traffic sign detection a challenging task are even more of a broad challenge for the text in road signs. It is similarly evident that text detection in scenes is itself challenging, and text-based traffic signs also falls into this category.

Color and shape are two distinguishing features of traffic signs, based upon which multiple algorithms are developed for their detection. These algorithms are categorized as shape, color or hybrid means combination of both color and shape based approaches [4]. Recently, these methods are divided into two main categories. First one is machine learning or deep learning methods and second one is based on other previous categories as traditional [26, 27]. Traditional detection methods are developed by using specific color spaces like RGB[28, 29], RGBN[28], HSV[30], and YCbCr [31] etc., or by using shape characteristic like Hough transform, or combination of both features. Machine learning-based methods [32] are mostly deployed using techniques such as support vector machine (SVM), Adaboost or random forest, and some other methods as well. Deep learning-based methods are progressing day by day, but the limitation of dataset is still there to achieve high accuray. These methods require massive amount of data or feature sets along with specialized hardware requirements, which make them time consuming, complex and costly during training.

B. SCENE TEXT RECOGNITION AND TEXT BASED TRAFFIC SIGNS

Traffic signs encompass symbol, specific text or both. The text within traffic signs provides semantic information that is helpful for drivers while driving. Text detection in traffic signs is classified under text recognition in scenes and come across similar problems as those encountered in scene text detection. Text detection and recognition in scenes is a vigorous and challenging topic for researchers working in computer vision field [33] due to vibrant nature of text i.e., blur, occlusion, noise, non-uniform illumination, clutter background, perspective distortion, multi-orientation, multi-lingual, font and much more in text. Despite a considerable amount of work in this area, ample space for improvement and enhancement is there to make robust and efficient practical systems [34].

Communication systems such as car navigation systems are taking advantage of text recognition systems, which prove to be quite helpful and beneficial. There are millions of people who are either fully or partially visually impaired in accordance to the report by the World Health Organization (WHO) [35] and such applications are highly helpful for them. Text recognition systems typically consist of text extraction, text classification, and text recognition. In the text extraction stage, potential candidates from natural scene images are taken out through performing certain enhancement, detection and segmentation approaches. Subsequently, the extracted potential candidates are further grouped in two classes of text region and non-text region. The identified text then recognizes in three phases: initial character by character, then word by word and last of all end to end.

Multiple state-of-the-art algorithms have been developed to date for natural scene text understanding as compared to text detection in traffic signs. These methods for text understanding serve as an inspiration for textual road signs detection and recognition of text in them. Text detection methods may be classified based on different approaches, including sliding window approach, texture based approach, connected component based approach, hybrid approach, and Convolution Neural Network (CNN) or deep learning based methods [36].

Connected component, also called region, based methods adopt a bottom-up strategy in which potential candidates are extracted and filtered out based on some predefined rules. These methods are usually scale invariant and provide character segmentation. Though, they are sensitive to noise and may suffer from declined efficiency for low contrast images. Statistical models are one of the way to overcome such issues [37].

Dense or opaque text may be treated as texture component [38]. Therefore, texture features are used in texture-based text detection methods to categorize the regions of interest (RoIs) into two classes (text or non-text regions). Such methods take text into account as a distinct texture component. Wavelet coefficient, filter responses, HOG and LBP, Fourier Transform (FT), Discrete Fourier

Transform (DFT) are the most commonly used texture features in these methods, often in conjunction with a multi stage sliding window.

Hybrid methods merge texture-based method and connected component methods, combining their advantages. Hybrid methods are paramount to boost text localization [39-41].

Currently, the advent of deep learning, with its astonishing characteristics, has significantly elevated the success rate across various domain of computer vision, and enrich the methods for classification to recognition and other related tasks. In deep learning, CNN is usually used for vision associated tasks, while Recurrent Neural Networks (RNN) are commonly involved in natural language processing (NLP) and tasks that are related to machine translation [42, 43].

Moreover, the researchers become able to take localization and recognition accuracy to a new high by leveraging big data concerning text extraction for training, after preeminence of deep learning in the field [21, 44-46].

Earlier, Wu et al.[47] proposed an algorithm to specifically detect text in textual traffic signs from their private dataset of videos by using k means, Gaussian Mixture Models (GMM) and geometry alignment analysis at 89% detection rate.

Reina et al.[48] developed a method using color segmentation with Fast Fourier Transform (FFT) for white and blue regions extraction and Optical Character Recognition (OCR) for character detection by removing non character regions. In between, regions relocation is performed to line up text with horizontal alignment and histogram analysis is for segmentation of text regions.

Rong at al.[12] proposed a method based on Cascaded Localization Network (CLN) with two customized convolutional nets, to automatically detect panels and text on them. This approach utilized string-wise text region detection instead of character-wise text saliency detection. Additionally, a temporal fusion is also deployed to reduce false positive errors. Moreover, a deep recurrent model, exclusive of character segmentation, is incorporated for text region recognition. The new top-down framework is trained and evaluated on a new dataset of panels i.e., Traffic Guide Panel. The paper lacks recognition results but detection with a precision rate of 73% and recall rate of 64 %.

Zhu et al. [17] proposed an algorithm for text detection in multilingual text based traffic signs in which a fully convolutional network (FCN) is used to extract traffic signs from images in Traffic Guide Panel dataset and then a word-based text detector, inspired from TextBoxes [20], using fast neural network(NN) is used for text detection in the panels with a precision and recall rate of 0.90 and 0.87 respectively. However, this method could not acquire complete text lines.

Peng et al. [14] proposed a deep learning based cascade detection model for text detection in traffic signs, utilizing MobileNetV2 for optimization. Modified version of SSD (Single Shot MultiBox Detector) network is used to locate text regions and text is extracted using a rotation-based text detection network. The model is evaluated on different datasets and provide a precision of 93% and recall of 91%.

Saba et al. [49] devised an approach using tiny YOLOv3 network for text panel detection on a novel dataset and obtained 97.3% precision, 94.5% recall rate and 95.5% F1 score by using k-fold cross validation method.

Altogether the state-of-the-art methods work well for images in certain type of environment and most of them work to the text localization stage except a few [50, 51]. The proposed method has been developed to detect panels and text localization in all environment by overcoming all the obstacles along with detected text recognition. The endeavor has been made to model a robust approach for text detection and text recognition to craft a wide-ranging methodology.

IV. PROPOSED METHODOLOGY

The proposed work focuses on extracting and understanding the text within the traffic sign on roads that have some kind of information. The aim of the proposed methodology is to accomplish the task with reduced false positive and good accuracy in a robust manner. For this purpose, we designed a semi pipelined method in which YOLOv5 is extended to extract traffic signs containing text and text panels from natural road text scene in an uncontrolled environment. Later the extracted text regions are preprocessed before its localization of text with the help of Fast Fourier Transform and Inverse Fast Fourier Transform (FFT_IFFT) filtration. LUV color space is used to get L band from the enhanced regions. The regions are transformed into the LUV color space first, and then only L color band, due to better edges, is extracted for next stages. Then, Maximally Stable Extremal Regions (MSER) incorporated for localization of text in detected panels. Subsequently, Optical Character Recognition (OCR) used for text recognition. Lastly, text correction process is performed through Natural Language Processing (NLP). Although, deep learning methods may perform text detection better than any other method but their call for enormous extent of parameter settings caused memory shortage with plummeting detection speed. FIGURE 2 presents an overview of proposed methodology.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3401044

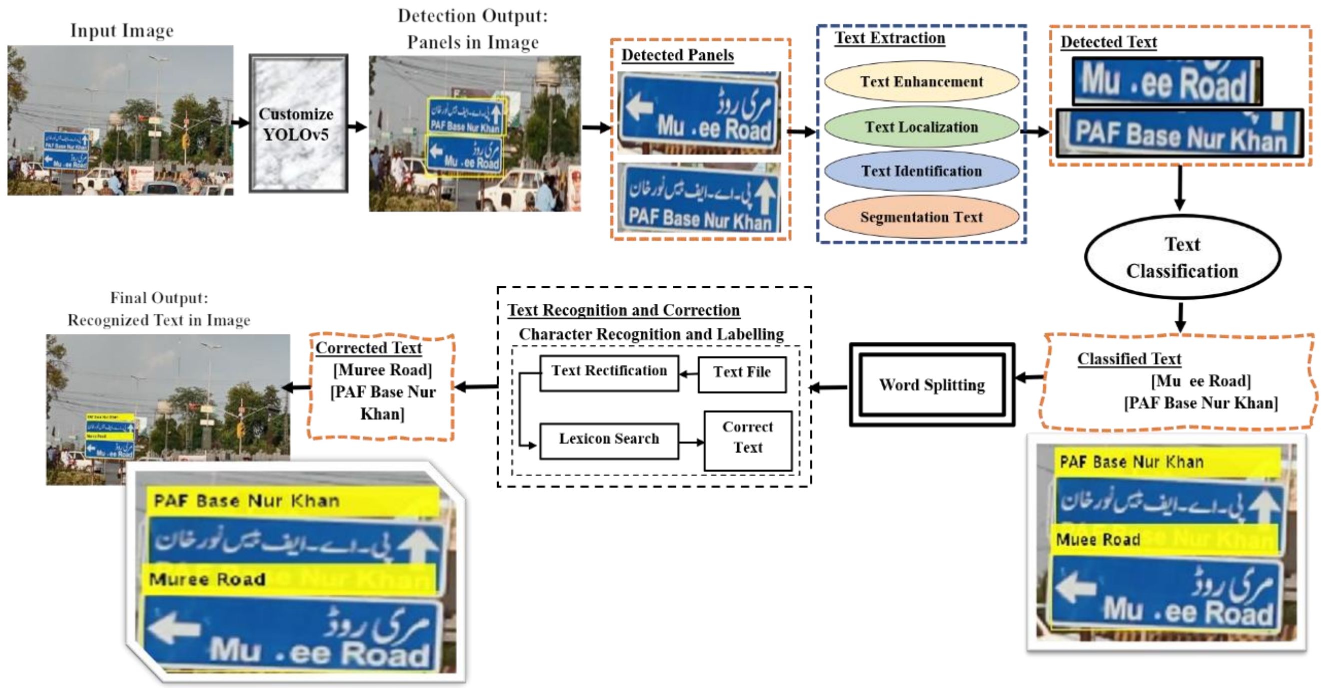


FIGURE 2. **Proposed Methodology for Text-based Traffic Signs Detection and Recognition**

A. TRAFFIC ROAD SIGN BOARD LOCALIZATION

Traffic road sign board localization from natural road scene is a complex task due to several obstructions like similar objects, occlusion, weather, illumination and so on. Researchers have developed many algorithms to overcome these obstacles but still have a lot to improve. Although with the upsurge of deep learning techniques these hurdles become a bit easier to deal with. In this work, we customized YOLOv5 to extract road signs from captured or acquired image. After extracting road signs from input image, they are classified into text and non-text traffic signs by YOLOv5, as shown in FIGURE 3, to further process them according to their needs or challenges.

YOLO is a single stage detector composed of a fully convolution NN that uses CSPDarkenet53 with a Focus Structure that trim down layers with CUDA memory requirement and enhanced forward and backward propagation [52] as a backbone feature extractor, with PANet as neck and YOLO layer as head. Binary cross entropy is used in YOLOv5 for loss function, along with logits loss function.

Either images or videos are used as input to be fed for YOLOv5, upon which likewise YOLOv4 preprocessing with mosaic data augmentation [53] and adaptive image filling is performed before sent to backbone network. Later, Focus Structure uses slicing operation to compress image data and two CSP branches i.e., CSP1_X and CSP2_X are employed for backbone network to extract features. Path Aggregation Network (PA Net) is a bottom to top feature pyramid uses at neck to progress information flow and localization in lower layers, which lead to upsurge in the localization accuracy, along with a top to bottom Feature Pyramid Network (FPN) that uses unsampling operation with transmitting and fusing information for acquiring predicted feature maps [54]. Head in YOLOv5 generates three diverse feature maps that are similar to YOLOv3 and Yolov4. The purpose of these maps is to attain predictions of multiscale. But the loss function differs from the rest.

While training YOLOv5 small for our problem, preprocessed images are resized to 640 × 640 before sending to the network. The backbone of the YOLOv5 is made up of five convolutional layers (CLs) with four convolutional blocks (CBs) along with a pooling layer. The head of network constitute of 4 six-layers blocks that contain one convolutional layer with one up sample layer and a concatenation layer with a block of three CLs. Two CLs, one convolutional block (three CLs), one CLs, one convolutional block (three CLs), one CLs, one convolutional block (three CLs), one CLs, one convolutional block (three CLs) and a polling layer make up the backbone of the network. The head consists of a block of six layers (one convolution, one up sample layer, one concatenation layer, three CLs block), six layered block (one convolution, one up sample layer, one concatenation layer, three CLs block), six layers based block (one convolution, one up sample layer, one concatenation layer, three CLs block), six layered block (one convolution, one up sample layer, one concatenation layer, three CLs block) and at the end of network a detection layer is formed.

The loss function at head in YOLOv5 is Generalized IoU Loss i.e., GIoU Loss, that trouble shoot the issue of non-coinciding bounding boxes effectively [55, 56]. Let L_{GIoU} represents GIoU Loss, so it can be defined as:

$$L_{GIoU} = 1 - GIoU \quad (1)$$

Where $GIoU$ is Generalized IoU that can be defined as:

$$GIoU = IoU - \frac{|C - (BB \cup BB_{GT})|}{|C|} \quad (2)$$

Here IoU is Intersection over Union that is used for finding out about predicted and ground truth bounding box boundary ratio. Whereas the $BB(x, y, w, h)$ and $BB_{GT}(x_{GT}, y_{GT}, w_{GT}, h_{GT})$ represent bounding boxes of predicted and ground truth box, and C represents the smallest box among the two bounding boxes BB and BB_{GT}. Similar way, IoU is expressed as follow:

$$IoU = \frac{|(BB \cap BB_{GT})|}{|(BB \cup BB_{GT})|} \quad (3)$$

After calculating GIoU, GIoU_Loss is calculated by subtracting C from IoU of BB and BB_{GT}.

$$GIoU_{Loss} = 1 - GIoU = 1 - IoU - \frac{|C - (BB \cup BB_{GT})|}{|C|} \quad (4)$$

FIGURE 3. Traffic Panel Detection Using YOLOv5s in Images of Different Orientation and Resolutions: (a) and (b) Portrait Orientation and (c) and (d) Landscape Orientation

B. TEXT LOCALIZATION AND EXTRACTION IN ROAD SIGNS

Preprocessing plays vital role for enhancing the detection of text within detected traffic panels and facilitates the localization of text in the image. Preprocessing stage removes many obstacles such as noise, ensuring more precise detection. Subsequently, to localize text in image MSER is applied, which is highly useful and widely adopted approach for text detection.

C. ENHANCEMENT IN LOW CONTRAST IMAGES

To enhance the extracted traffic sign panels, noise removal using Gaussian filter is initially applied. Later, for low contrast text images, FFT-IFFT is applied to get the enhanced image. Let $I(i, j)$ is an input image with a dimension of $m \times n$ that contains text in it. The FFT of an image $I(i, j)$ can be defined as:

$$F(x, y) = \sum_{i=0}^{n-1} \sum_{j=0}^{m-1} I(i, j)\ e^{-k2\pi\left(\frac{xi}{n} + \frac{yj}{m}\right)} \quad (5)$$

where $F(x, y)$ is image after applying FFT, $k = \sqrt{-1}$, x and $i = 0,1,2, \dots, n$; y and $j = 0,1,2,3, \dots, m$. The filtration function $LT\,(x, y)$ is logarithmic transformation. The filtration function is used to plots low input gray scale values of narrow range to output values of wider range. The filtering function $LT\,(x, y)$ can be described as:

$$LT(x, y) = \mathrm{K} * \log\left(1 + \mathrm{I}\,(i, j)\right) \quad (6)$$

here K is a constant value that is normally considered 1 for enhancement operation and $I(i, j)$ is an image pixel to be processed.

After this, IFFT is applied to get back the enhanced image in spatial domain from frequency domain. IFFT can be elaborated as:

$$I(i, j) = \frac{1}{nm} \sum_{x=0}^{n-1} \sum_{y=0}^{m-1} F(x, y) e^{k2\pi\left(\frac{xi}{n} + \frac{yj}{m}\right)} \quad (7)$$

The image is further transformed to LUV color space as described in [57]. To covert the RGB image $I(i, j)$ to LUV image $L(i, j)$, initially the RGB image $I(i, j)$ is converted to CIE XYZ image $I_{XYZ}(i, j)$, where X, Y, Z are tri-stimulus values of CIE, using following formula:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} 0{:}412453 & 0{:}357580 & 0{:}180423 \\ 0{:}212671 & 0{:}715160 & 0{:}072169 \\ 0{:}019334 & 0{:}119193 & 0{:}950227 \end{bmatrix} * \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (8)$$

After RGB to CIE XYZ conversion, image $I_{XYZ}(i, j)$ is converted to LUV image $L(i, j)$ with the help of following equations:

$$L = \begin{cases} C1\left(\frac{Y}{Y_n}\right)^{\frac{1}{3}} - c, & if\ \frac{Y}{Y_n} > T \\ C2\left(\frac{Y}{Y_n}\right) - c, & if\ \frac{Y}{Y_n} \leq T \end{cases} \quad (9)$$

$$U = kL * (\bar{u} - \bar{u}_n) \quad (10)$$

$$V = kL * (\bar{v} - \bar{v}_n) \quad (11)$$

Where Y_n is luminance of a white object; $C1, C2$ and T are constants and their values are

$$116, \left(\frac{29}{3}\right)^3 and \left(\frac{6}{29}\right)^3$$

respectively; c and k are constant values i.e., 16 and 13 correspondingly; and $\bar{u}_n, and\ \bar{v}_n$are chromaticity coordinates of $(\bar{u}, \bar{v})$ for white object and following equations are used to calculate $\bar{u}, \bar{v}, \bar{u}_n, and\ \bar{v}_n$.

$$\bar{u} = \frac{4X}{X + 15Y + 3Z} \quad (12)$$

$$\bar{v} = \frac{9Y}{X + 15Y + 3Z} \quad (13)$$

$$\bar{u}_n = \frac{4X_n}{X_n + 15Y_n + 3Z_n} \quad (14)$$

$$\bar{v}_n = \frac{9Y_n}{X_n + 15Y_n + 3Z_n} \quad (15)$$

The reason to transform text image into LUV is that it provides adjusted illumination and gradient with improved region level contrast and local stability which are very helpful in text localization. LUV is computationally faster than other color spaces as well. Its L channel is better to work on text than U channel as it shows suppressed edges.

D. TEXT LOCALIZATION USING MSER

After enhancing the region of interest (RoI), text localization is performed. Text localization means to identify text and is done through Maximally Stable Extremal Regions (MSER) [58]. MSER is a connected component-based approach using watershed principle for segmenting the regions. It is a paramount, well known and robust method [59] for multi sized and multi oriented text detection, as it is covariant to adjacency preserving and rotation, affine and scale invariant but blur. Usually, text-based road signs are designed to make text readability easier rather than being attractive, so they are mostly in contrast with their background color, like lighter color background with duskier color for text and vice versa. This attribute of text-based road signs makes MSER an appropriate choice for text detection. MSER also not computationally costly rather, it is efficient, exhibiting complexity that is nearly linear [61]. For real time applications, linear complexity works as an exceptionally significant attribute. Regions detected using MSER method are more effective in connected component analysis as they not only provide geometric features but color uniformity. Thus, in proposed method, MSER is applied on L channel of image $L(i,j)$ due to more clear edges in L channel. The process can be defined as:

$$LCC_{MSER}[n] = extractMSER(L) \quad (16)$$

where $LCC_{MSER}[n]$ represents n number of connected components extracted via $extractMSER$ function by passing L channel as an argument.

Afterwards, some geometric properties like Aspect ratio, Eccentricity, Euler number, Extent, Solidity are manipulated to eliminate non-text bobs detected and present in $LCC_{MSER}[n]$. It can be defined as:

$$LCC_{MSER]}[n] = \begin{cases} 1 & if(ec > .995\ \&\ s < .3\ \&\ eu < -4\ \&\ .9 < ex < .2\ \&\ ar > 3) \\ 0 & otherwise \end{cases} \quad (17)$$

where ec, s, eu, ex and ar denotes Eccentricity, Solidity, Euler number, Extent and Aspect ratio respectively. Here eccentricity, ec can be defined as ratio between minor axis symbolized as ax_{min} and major axis represented via ax_{max} of the object and formulate as:

$$ec = \frac{ax_{min}}{ax_{max}} \quad (18)$$

The solidity calculates the density of an object with the help of area of the object, which is denoted by a and convex area represented by ca of that object. The solidity s can be defined by using following equation:

$$s = \frac{a}{ca} \quad (19)$$

Euler number eu is basically a difference between connected components CC and holes H in an image. The equation used to calculate Euler number, eu:

$$eu = CC - H \quad (20)$$

The extent ex is a ratio between area of the object i.e., a and area of bounding box denoted by a_{bb}, which can be written in equation form as follows:

$$ex = \frac{a}{a_{bb}} \quad (21)$$

The aspect ratio ar is expressed in terms of a ratio between length l and width w i.e.,

$$ar = \frac{l}{w} \quad (22)$$

E. TEXT RECOGNITION AND CORRECTION

The filtered-out regions are further processed to recognize text using OCR. Later, for correction phase, a text file is created by putting away all the acknowledged characters from previous stage.

The output received from previous stage in text file consist of recognized character by OCR. Regular OCR may detect incorrect character or may fail to detect some character. A correction phase is designed to encounter the issue of incorrect or missed detection by OCR. The correction phase is designed using hamming distance to avoid any inversion of character in text recognized by OCR.

Hamming distance is used for correcting the text meaning on traffic sign board by calculating total number of bits/characters altered in observed string i.e., character stored in text file. Hamming distance can be defined as:

$$HD = count\ (s_1 \oplus s_2) \quad (23)$$

where s_1 and s_2 represent binary strings of same length on which XOR operation is applied and then passed to $count$ function that return total number of 1's in string i.e., hamming distance denoted by HD. The value of hamming distance talks about unrecognized character.

String containing character labels that are stored in text file are additional processed using lexicon. The string is searched using lexicon where hamming distance is calculated, based on which string correction is performed as zero distance means same string with no variation, smaller distance means more similarity and larger distance represents great dissimilarity. If the distance between the text in sign board and text in text file is zero, then no correction is required otherwise string correction is performed to get right outcome. In case of non-zero distance between strings, the faulty string is again searched using lexicon to find some list of words that have less distance and possibly one of them is the actual word.

Let RC is the recognized string produced by OCR, upon which processing is performed to get the right outcome $\overline{RC}$ using following equation for correction:

$$\overline{RC} = \begin{cases} RC & if\ HD = 0 \\ Change(RC) & otherwise \end{cases} \qquad (24)$$

where hamming distance HD is calculated and if found $0\ (zero)$, no further processing is performed instead on non-zero HD, a function $Change$ is applied on RC to perform required correction. The end result is depicted in FIGURE 4.

FIGURE 4. Text Recognition in detected Traffic Panel Detection

Algorithm

Step 1: Input image I → Trained YOLOv5()

Return guided signs $I_g(i)$ where i=1,2, …

Step 2: Text Enhancement in Low Contrast Images

<u>Step 2a:</u> FFT_IIFT filtration

$I_{gf}(i)$ ← FFT($I_g(i)$)

$I_{gff}(i)$ ← logTransFilter($I_{gf}(i)$)

$I_{gffi}(i)$ ← IFFT($I_{gff}(i)$)

<u>Step2b:</u> RGB image into LUV image and L channel extraction

$I_{LUV}(i)$ ← LUV($I_{gffi}(i)$)

$L(i)$ ← extract ($I_{LUV}(i)$)

Step 4: Text Localization using MSER

<u>Step 4a:</u> $LCC_{MSER}[n] = extractMSER(L)$

where n= number of connected components

<u>Step 4b:</u> $LCC_{MSER]}[n]$ further filtered as:

$$LCC_{MSER]}[n] = \begin{cases} 1, & if\,(A\ \&\ B\ \&\ C\ \&\ D\ \&\ E\,) \\ 0, & otherwise \end{cases}$$

where A, B, C, D and E are conditions defined as following:

$A =\ ec > .995$

$B =\ s < .3$

$C = eu < -4$

$D =\ .9 < ex < .2$

$E = ar > 3$

Step 5: Text Recognition and Correction

RC[i] ← OCR($LCC_{MSER]}[n]$)

$HD\ =\ count\ (s_1 \oplus\ s_2)$

$$\overline{RC} = \begin{cases} RC & if\ HD = 0 \\ Change(RC) & otherwise \end{cases}$$

Step 6: Recognized Text in Guided Sign → $\overline{RC}$

V. RESULTS

The proposed work is used to locate and recognize text in road signs, which are essential for guidance and safety purposes. Robotically detecting and recognizing of text in road signs is obliging for individuals on and around the road, along with vehicles. This text detection in road signs aids in evading waste of time and inability to see those sign due to certain reasons. We have performed multiple experiments in different settings with different testing data, and all outcomes from those experiments are better and satisfactory.

Initially, Yolov5s is deployed for detecting traffic signs containing text and guided panels which yields very high-quality outcomes, preceded by a full fledge training on a novel polyglot and adaptable dataset named ASAYAR [60]. This dataset contains guided signs, regulatory signs and warning signs. However, for our work we put all signs under two categories that are guided and unguided traffic signs. The results of YOLOv5s in comparison with other models [60] is displayed in TABLE 1. The table clearly indicates that the YOLOv5 outperformed other models by detecting almost all guided signs.

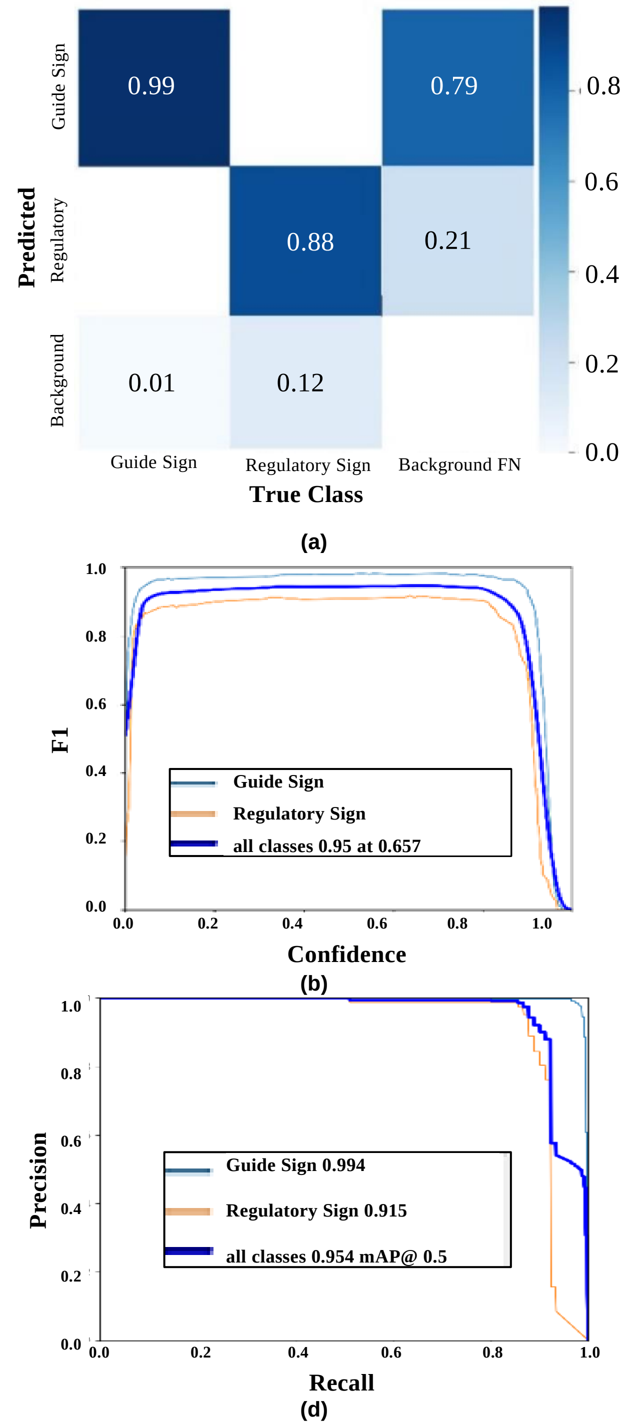
TABLE 1. Panel Detection Outcomes on Different Models presented in [60] and Customized YOLOv5 of Proposed Work

Model	Class	AP	mAP@0.5
Faster RCNN	Guided	95.47	76.27
	Warning	59.11	
	Regulatory	74.23	
Retina net	Guided	94.81	75.08
	Warning	58.82	
	Regulatory	71.61	
SSD	Guided	93.72	72.45
	Warning	55.11	
	Regulatory	68.53	
Faster RCNN + Base Line	Guided	97.32	78.85
	Warning	63.10	
	Regulatory	77.13	
YOLOv5s	Guided	99.40	95.40
	Unguided	91.50	

YOLOv5 is trained on different epochs like 20, 50, and 100. It shows better results at 100 epochs. The labels are

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3401044

divided into two categories guided and regulatory traffic signs where warning signs are counted as regulatory ones. The graphical results of Yolov5s detection are given in FIGURE 5 that illustrate different evaluations like confusion matrix, confidence against recall, precision and F1 score. All of them presenting a better performance of YOLOv5.



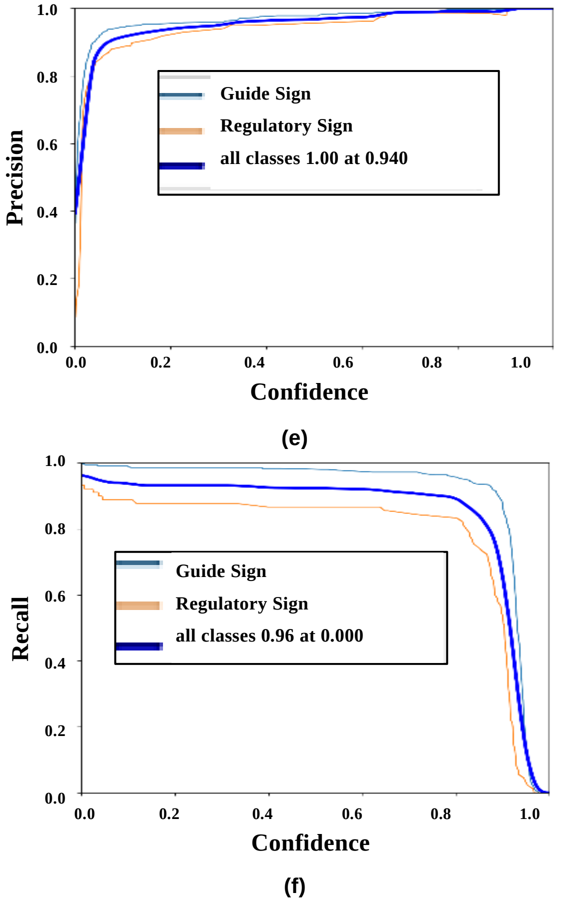


FIGURE 5. Different Metrics of Trained YOLOv5s Model for Detection (a) Confusion Matrix; (b) Confidence Vs F1 Score Curve; (c) Correlogram; (d) mAP@0.5; (e) Confidence Vs Precision Curve;(f) Confidence Vs Recall Curve

After panel detection, text localization is done via MSER after applying enhancement which results in better outcomes than simple MSER. The approach is computationally effective and successfully localize most of the text in panels.

The comparison of proposed technique for detection with other existing approaches in relations to time, recall, precision, and F1/$F_{measurre}$ score is depicted. It can be seen from the tables, that the proposed work technique performs much better than other existing techniques for panel detection. Although there is ample space for improvement especially regarding text detection.

The result of panel detection can also be seen from FIGURE 6. The comparison in the graphs clearly exhibiting the better performance of proposed algorithm by using YOLOv5.

We also perform testing using different datasets [17, 60-62] for detection. We also collect images using mobile phones, camera and internet, which we have transformed into an unruffled dataset. The detection rate for almost all dataset is quite satisfactory.

TABLE 2. Relationship/ Contrast with SOTA Detection Techniques of Road Signs/Panels containing Text

Year	Author & Method	Dataset	Recall	Precision	$F_{measure}$	Time
2018	Zhu et al. [8]: FCN	TTSDCE & Traffic Guide Panel dataset	0.935	0.941	0.938	0.14s
2020	Peng et al. [43]: improved SSD	Traffic Guide Panel, TT 100K & self-collected dataset.	0.942	0.947	0.944	0.26s
2022	Saba et al. [49] Tiny YOLOv3	Persian text-based traffic panels	0.970	0.94	0.95	—
Proposed Method: customized YOLOv5s		ASAYAR & Self collected dataset	0.961	0.99	0.95	0.12s

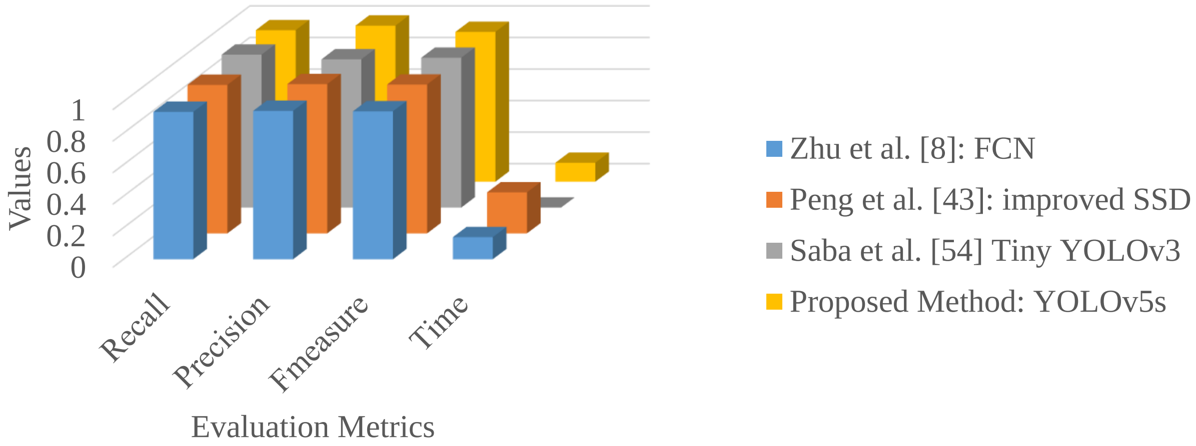


FIGURE 6. Comparison with SOTA Techniques for Detecting Road Signs/Panels

Although due to some issue like occluded panel, far off panel, too much shadow etc., the detection falls through a bit. It is because of less samples while training. To increase the robustness of YOLOv5s more samples of such conditions are required. The performance of proposed method is shown in TABLE 3, which clearly exhibiting a decent outcome on almost all datasets.

TABLE 3. Panel Detection Testing Result on different Datasets

Dataset	Recall	Precision	F1
Traffic Guide Panel Dataset [16]	0.956	0.986	0.950
ASAYAR [60]	0.961	0.990	0.945
Self-Collected Dataset	0.950	0.982	0.940
ATTICA [61]	0.944	0.976	0.933
ATSD [62]	0.915	0.923	0.919

The result of YOLOv5 panel detection on different dataset is presented in FIGURE 7 and FIGURE 8, which shows robustness and precision of customized YOLOv5. The detection rate for almost all dataset is quite satisfactory, although due to some issue like occluded panel, far off panel, too much shadow etc. the detection falls through a bit, because of less samples while training. To increase the robustness of YOLOv5s more samples of such conditions are required.

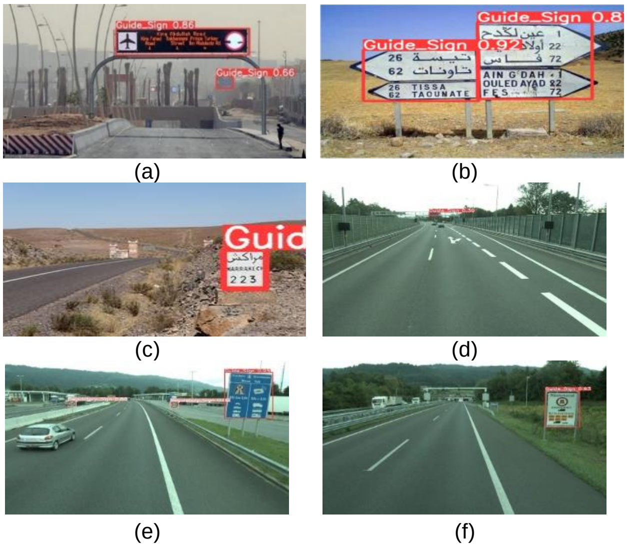
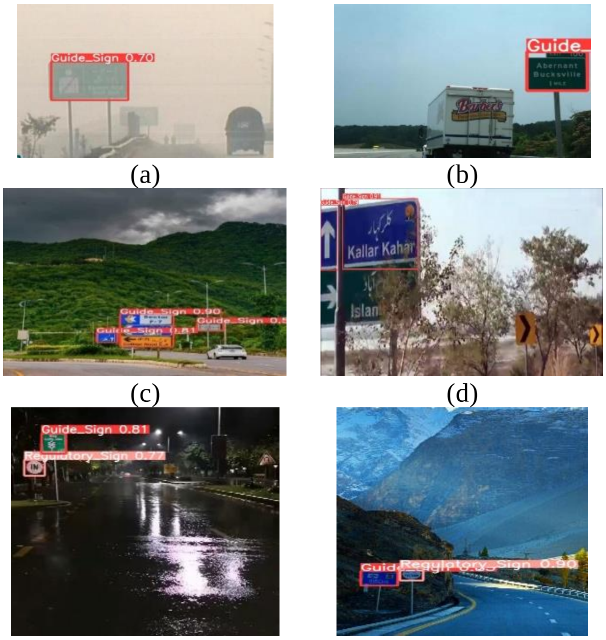


FIGURE 7. Panel Detection on ATTICA (a - c) and ATSD (d – f) Dataset

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal



(a) (b) (c) (d) (e) (f)

FIGURE 8. Panel Detection on Self-Collected Dataset

After panel detection, text localization is done via MSER after applying enhancement which results in better outcomes than simple MSER. The approach is computationally effective and successfully localizes most of the text in panels.

in terms of all measures by using customized YOLOv5 for detection, which is trained extensively on GPU using ASAYAR dataset and tested on multiple datasets. The false positive rate is quite low, which makes the enhanced MSER work easier. MSER with enhancement performed really well on detected panels.

The result of text localization is also presented in FIGURE 9. The comparison in the graphs clearly exhibiting the better performance of proposed algorithm by using YOLOv5 and enhanced MSER.

TABLE 4 is depicting the comparison of proposed technique regrading text detection in traffic signs with other existing approaches. It can be seen from the table, that the proposed work technique performs much better than other existing techniques for text localization. The proposed technique beats other existing methods in terms of all measures by using customized YOLOv5 for detection, which is trained extensively on GPU using ASAYAR dataset and tested on multiple datasets. The

false positive rate is quite low, which makes the enhanced MSER work easier. MSER with enhancement performed really well on detected panels.

The result of text localization is also presented in FIGURE 9. The comparison in the graphs clearly exhibiting the better performance of proposed algorithm by using YOLOv5 and enhanced MSER.

TABLE 4. Comparison with SOTA Techniques for Text Detection in Road Signs/Panels

Year	Author and Method	Datasets	Recall	Precision	$F_{measure}$	Time
2016	Rong et al. [9]: CLN + Deep recurrent Model	Traffic Guide Panel dataset	0.730	0.640	0.680	.16s/167ms
2018	Zhu et al. [8]: TextBox (longer convolutional kernels)	TTSDCE & Traffic Guide Panel dataset	0.900	0.870	0.880	.15s/154ms
2020	Peng et al. [43]: improved EAST (MobileNetV2)	Traffic Guide Panel, TT 100K & self-collected dataset.	0.930	0.910	0.920	.12s/120ms
Proposed Method		ASAYAR & Self collected dataset	0.945	0.922	0.930	.10s/104ms

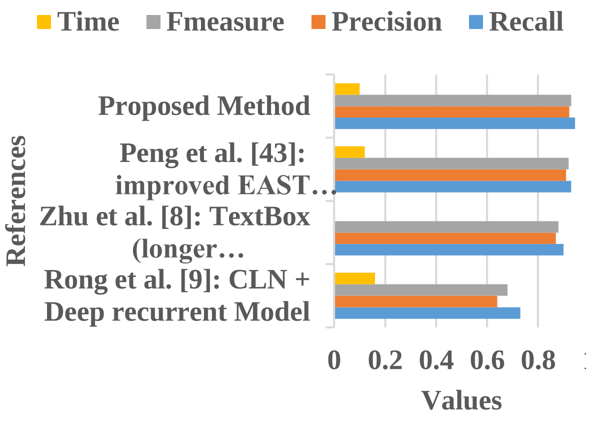


FIGURE 9. Comparison with State of Art Techniques for Text Detection in Road Signs/Panels

The proposed approach also worked on recognition of the detected text which outcomes in statistical manner are presented in TABLE 5. A few researchers have worked on this stage to recognize detected text. In TABLE 5, the proposed method results for recognition are compared with state-of-the-art methods, and it is evident the proposed work done a quite satisfactory job with help of OCR and NLP.

TABLE 5. Comparison Results for the Recognition of Text Stage

Method	Precision	Recall	F1
Xavier et al. [50]	0.96	0.960	0.952
Greenlagh et al. [15]	0.87	0.910	0.890
Jain et al.[51]	0.96	0.931	0.930
Proposed System: OCR and NLP	0.98	0.983	0.965

In TABLE 6, is constructed using different types of OCR approaches presented in [16] and proposed approach of OCR with NLP which clearly demonstrate the enhanced performance of OCR with NLP.

TABLE 6. Recognition Rate Comparison between Different OCRs

OCR Names	Precision	Recall	F 1
Standard Tesseract OCR [15]	0.48	0.34	0.4
OCR with shape correction [15]	0.69	0.75	0.72
OCR with temporal fusion [15]	0.83	0.33	0.47
Proposed	0.98	0.983	0.965

Moreover, different OCRs performance on detected text with and without NLP is also computed in our work which is mounted in TABLE 7, which indicates that using NLP makes the performance and outcomes better.

TABLE 7. Comparison of OCRs performances with and without NLP using F1 Score

OCR Names	F1 without NLP	F1 with NLP
Tesseract OCR	0.4	0.55
OCR with shape correction	0.72	0.841
OCR with temporal fusion	0.47	0.61

VI. CONCLUSION

The proposed method based on three stages; first stage is designed to detect guided and unguided panels using YOLOv5s; in second stage preprocessing is done before giving the detected guided panels to MSER for localization of text; in the last stage of recognition OCR with NLP is used to recognize, refine and correct the text in panels. The YOLOv5s performance is almost outperformed other methods and due to preprocessing step MSER also performed well. Later NLP with OCR enhanced the recognition rate and beat many state-of-the-art algorithms. Although, the performance of the proposed system has improved but still there is ample space for improvement is there such as text in haze, and text in cluttered background.

ACKNOWLEDGEMENT

This study is supported via funding from Prince sattam bin Abdulaziz University project number (PSAU/2024/R/1445)

REFERENCES

[1] Q. Wang, J. Wan, and Y. Yuan, "Locality constraint distance metric learning for traffic congestion detection," *Pattern Recognition*, vol. 75, pp. 272-281, 2018.
[2] E. Wang, Y. Li, A. Sun, H. Gao, J. Yang, and Z. Fang, "Road detection based on illuminant invariance and quadratic estimation," *Optik*, vol. 185, pp. 672-684, 2019.
[3] J. M. Á. Alvarez and A. M. Ĺopez, "Road detection based on illuminant invariance," *IEEE transactions on intelligent transportation systems*, vol. 12, pp. 184-193, 2010.
[4] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, pp. 1484-1497, 2012.
[5] A. Gudigar, S. Chokkadi, and U. Raghavendra, "A review on automatic detection and recognition of traffic sign," *Multimedia Tools and Applications*, vol. 75, pp. 333-364, 2016.
[6] Y. Saadna and A. Behloul, "An overview of traffic sign detection and classification methods,"

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3401044

International journal of multimedia information retrieval, vol. 6, pp. 193-210, 2017.

[7] S. B. Wali, M. A. Abdullah, M. A. Hannan, A. Hussain, S. A. Samad, P. J. Ker, and M. B. Mansor, "Vision-based traffic sign detection and recognition systems: Current trends and challenges," *Sensors*, vol. 19, p. 2093, 2019.

[8] B. Sanyal, R. K. Mohapatra, and R. Dash, "Traffic sign recognition: A survey," in *2020 International Conference on Artificial Intelligence and Signal Processing (AISP)*, 2020, pp. 1-6.

[9] J. Wang, Y. Chen, Z. Dong, and M. Gao, "Improved YOLOv5 network for real-time multi-scale traffic sign detection," *Neural Computing and Applications*, vol. 35, pp. 7853-7865, 2023.

[10] A. González, L. M. Bergasa, J. J. Yebes, and M. Sotelo, "Automatic information recognition of traffic panels using SIFT descriptors and HMMS," in *13th International IEEE Conference on Intelligent Transportation Systems*, 2010, pp. 1289-1294.

[11] Á. González, L. M. Bergasa, J. J. Yebes, and J. Almazán, "Text recognition on traffic panels from street-level imagery," in *2012 IEEE Intelligent Vehicles Symposium*, 2012, pp. 340-345.

[12] X. Rong, C. Yi, and Y. Tian, "Recognizing text-based traffic guide panels with cascaded localization network," in *European Conference on Computer Vision*, 2016, pp. 109-121.

[13] G. H. de Oliveira, F. A. da Silva, D. R. Pereira, L. L. de Almeida, A. O. Artero, A. F. Bonora, and V. H. C. de Albuquerque, "Automatic detection and recognition of text-based traffic signs from images," *IEEE Latin America Transactions*, vol. 16, pp. 2947-2953, 2018.

[14] X. Peng, X. Chen, and C. Liu, "Real-time traffic sign text detection based on deep learning," in *IOP Conference Series: Materials Science and Engineering*, 2020, p. 072039.

[15] W. Sun, Y. Du, X. Zhang, and G. Zhang, "Detection and recognition of text traffic signs above the road," *International Journal of Sensor Networks*, vol. 35, pp. 69-78, 2021.

[16] J. Greenhalgh and M. Mirmehdi, "Recognizing text-based traffic signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, pp. 1360-1369, 2014.

[17] Y. Zhu, M. Liao, M. Yang, and W. Liu, "Cascaded segmentation-detection networks for text-based traffic sign detection," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, pp. 209-219, 2017.

[18] J. Guo, R. You, and L. Huang, "Mixed vertical-and-horizontal-text traffic sign detection and recognition for street-level scene," *IEEE Access*, vol. 8, pp. 69413-69425, 2020.

[19] Y. Liu, J. Peng, J.-H. Xue, Y. Chen, and Z.-H. Fu, "TSingNet: Scale-aware and context-rich feature learning for traffic sign detection and recognition in the wild," *Neurocomputing*, vol. 447, pp. 10-22, 2021.

[20] M. Liao, B. Shi, X. Bai, X. Wang, and W. Liu, "Textboxes: A fast text detector with a single deep neural network," in *Thirty-first AAAI conference on artificial intelligence*, 2017.

[21] Z. Tian, W. Huang, T. He, P. He, and Y. Qiao, "Detecting text in natural image with connectionist text proposal network," in *European conference on computer vision*, 2016, pp. 56-72.

[22] G. J. Ansari, J. H. Shah, M. Sharif, and S. ur Rehman, "A novel approach for scene text extraction from synthesized hazy natural images," *Pattern Analysis and Applications*, vol. 23, pp. 1305-1322, 2020.

[23] G. J. Ansari, J. H. Shah, M. Yasmin, M. Sharif, and S. L. Fernandes, "A novel machine learning approach for scene text extraction," *Future Generation Computer Systems*, vol. 87, pp. 328-340, 2018.

[24] S. K. Berkaya, H. Gunduz, O. Ozsen, C. Akinlar, and S. Gunal, "On circular traffic sign detection and recognition," *Expert Systems with Applications*, vol. 48, pp. 67-75, 2016.

[25] S. Khalid, J. H. Shah, M. Sharif, M. Rafiq, and G. S. Choi, "Traffic Sign Detection with Low Complexity for Intelligent Vehicles Based on Hybrid Features," *Computers, Materials & Continua*, vol. 76, 2023.

[26] S. Luo, L. Yu, Z. Bi, and Y. Li, "Traffic Sign Detection and Recognition for Intelligent Transportation Systems: A Survey," *Journal of Internet Technology*, vol. 21, pp. 1773-1784, 2020.

[27] A. Ellahyani, I. El Jaafari, and S. Charfi, "Traffic Sign Detection for Intelligent Transportation Systems: A Survey," in *E3S Web of Conferences*, 2021, p. 01006.

[28] S. Khalid, N. Muhammad, and M. Sharif, "Automatic measurement of the traffic sign with digital segmentation and recognition<? show

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3401044

[AQ="" ID=" Q1]"?," *IET Intelligent Transport Systems*, vol. 13, pp. 269-279, 2019.
[29] K. T. Phu and L. L. Oo, "RGB Color based Myanmar Traffic Sign Recognition System from Real-time Video," *International Journal of Information Technology (IJIT)*, vol. 4, 2018.
[30] A. Ellahyani, M. El Ansari, and I. El Jaafari, "Traffic sign detection and recognition based on random forests," *Applied Soft Computing*, vol. 46, pp. 805-815, 2016.
[31] H. Huang and L.-Y. Hou, "Speed limit sign detection based on Gaussian color model and template matching," in *2017 International Conference on Vision, Image and Signal Processing (ICVISP)*, 2017, pp. 118-122.
[32] S. H. A. Shah and J. H. Shah, "Road Scene Text Detection and Recognition Using Machine Learning," in *2023 IEEE 20th International Conference on Smart Communities: Improving Quality of Life using AI, Robotics and IoT (HONET)*, 2023, pp. 1-6.
[33] P. Naveen and M. Hassaballah, "Scene text detection using structured information and an end-to-end trainable generative adversarial networks," *Pattern Analysis and Applications*, vol. 27, p. 33, 2024.
[34] J. Ghosh, A. K. Talukdar, and K. K. Sarma, "A light-weight natural scene text detection and recognition system," *Multimedia Tools and Applications*, vol. 83, pp. 6651-6683, 2024.
[35] W. H. Organization. (2018). *Blindness and vision impairment*. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment
[36] D. Zhong, S. Lyu, P. Shivakumara, U. Pal, and Y. Lu, "Text proposals with location-awareness-attention network for arbitrarily shaped scene text detection and recognition," *Expert Systems with Applications*, vol. 205, p. 117564, 2022.
[37] H. I. Koo and D. H. Kim, "Scene text detection via connected component clustering and nontext filtering," *IEEE transactions on image processing*, vol. 22, pp. 2296-2305, 2013.
[38] J. Gllavata, R. Ewerth, and B. Freisleben, "Text detection in images based on unsupervised classification of high-frequency wavelet coefficients," in *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, 2004, pp. 425-428.
[39] C. Yi and Y. Tian, "Localizing text in scene images by boundary clustering, stroke segmentation, and string fragment classification," *IEEE Transactions on Image Processing*, vol. 21, pp. 4256-4268, 2012.
[40] C. Shi, B. Xiao, C. Wang, and Y. Zhang, "Graph-based background suppression for scene text detection," in *Document Analysis Systems (DAS), 2012 10th IAPR International Workshop on*, 2012, pp. 210-214.
[41] H. Chen, S. S. Tsai, G. Schroth, D. M. Chen, R. Grzeszczuk, and B. Girod, "Robust text detection in natural images with edge-enhanced maximally stable extremal regions," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 2609-2612.
[42] Y. Jiang, X. Zhu, X. Wang, S. Yang, W. Li, H. Wang, P. Fu, and Z. Luo, "R2cnn: Rotational region cnn for orientation robust scene text detection," *arXiv preprint arXiv:1706.09579*, 2017.
[43] W. Yu, L. Yin, C. Zhang, Y. Chen, and A. X. Liu, "Application of Quantum Recurrent Neural Network in Low Resource Language Text Classification," *IEEE Transactions on Quantum Engineering*, 2024.
[44] S. Tian, U. Bhattacharya, S. Lu, B. Su, Q. Wang, X. Wei, Y. Lu, and C. L. Tan, "Multilingual scene character recognition with co-occurrence of histogram of oriented gradients," *Pattern Recognition*, vol. 51, pp. 125-134, 2016.
[45] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, and X. Xue, "Arbitrary-oriented scene text detection via rotation proposals," *IEEE Transactions on Multimedia*, 2018.
[46] Y. Wang, C. Shi, B. Xiao, C. Wang, and C. Qi, "CRF based text detection for natural scene images using convolutional neural network and context information," *Neurocomputing*, vol. 295, pp. 46-58, 2018.
[47] W. Wu, X. Chen, and J. Yang, "Detection of text on road signs from video," *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, pp. 378-390, 2005.
[48] A. V. Reina, R. L. Sastre, S. L. Arroyo, and P. G. Jiménez, "Adaptive traffic road sign panels text extraction," in *Proceedings of the 5th WSEAS International Conference on Signal Processing, Robotics and Automation*, 2006, pp. 295-300.
[49] S. Kheirinejad, N. Riahi, and R. Azmi, "Text-Based Traffic Panels Detection using the Tiny YOLOv3 Algorithm."
[50] S. Xavier and R. Reshmi, "Automatic detection and recognition of text in traffic sign boards

based on word recognizer," *IJIRST–International Journal for Innovative Research in Science & Technology*, vol. 3, 2016.
[51] R. Jain and D. Gianchandani, "A hybrid approach for detection and recognition of traffic text sign using MSER and OCR," in *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2018 2nd International Conference on*, 2018, pp. 775-778.
[52] U. Nepal and H. Eslamiat, "Comparing YOLOv3, YOLOv4 and YOLOv5 for autonomous landing spot detection in faulty UAVs," *Sensors*, vol. 22, p. 464, 2022.
[53] C. Wu, W. Wen, T. Afzal, Y. Zhang, and Y. Chen, "A compact dnn: approaching googlenet-level accuracy of classification and domain adaptation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5668-5677.
[54] G. X. Hu, B. L. Hu, Z. Yang, L. Huang, and P. Li, "Pavement crack detection method based on deep learning models," *Wireless Communications and Mobile Computing*, vol. 2021, pp. 1-13, 2021.
[55] S. Li, X. Gu, X. Xu, D. Xu, T. Zhang, Z. Liu, and Q. Dong, "Detection of concealed cracks from ground penetrating radar images based on deep learning algorithm," *Construction and Building Materials*, vol. 273, p. 121949, 2021.
[56] H. Rezatofighi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 658-666.
[57] V. K. Shrivastava and M. K. Pradhan, "Rice plant disease classification using color features: a machine learning paradigm," *Journal of Plant Pathology*, vol. 103, pp. 17-26, 2021.
[58] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and vision computing*, vol. 22, pp. 761-767, 2004.
[59] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *International journal of computer vision*, vol. 65, pp. 43-72, 2005.
[60] M. Akallouch, K. S. Boujemaa, A. Bouhoute, K. Fardousse, and I. Berrada, "ASAYAR: A dataset for Arabic-Latin scene text localization in highway traffic panels," *IEEE Transactions on Intelligent Transportation Systems*, 2020.
[61] K. S. Boujemaa, M. Akallouch, I. Berrada, K. Fardousse, and A. Bouhoute, "ATTICA: A Dataset for Arabic Text-Based Traffic Panels Detection," *IEEE Access*, vol. 9, pp. 93937-93947, 2021.
[62] A. Maletzky, N. Hofer, S. Thumfart, K. Bruckmüller, and J. Kasper, "Traffic Sign Detection and Classification on the Austrian Highway Traffic Sign Data Set," *Data*, vol. 8, p. 16, 2023.

BIOGRAPHIES

SARA KHALID received the M.S. degree in computer science from COMSATS University Islamabad, Wah Campus, Pakistan. Currently she is pursuing her PhD degree from COMSATS. Since 2014, she has been in the education field as a teacher in Pakistan. Her areas of specialization are automation and pattern recognition. She has publications in IF, SCI, and ISI journals and in national and international conferences. Her research interests include deep learning, algorithms design and analysis, machine learning, image processing, and big data.

JAMAL HUSSAIN SHAH received the Ph.D. degree in pattern recognition from the University of Science and Technology of China, Hefei, China. Since 2008, he has been in the education field. He is currently an Associate Professor with COMSATS University Islamabad, Wah Campus, Pakistan. His areas of specialization are automation and pattern recognition. He has 60 publications

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2024.3401044

in IF, SCI, and ISI journals and in national and international conferences. His research interests include deep learning, algorithms design and analysis, machine learning, image processing, and big data.

MUHAMMAD SHARIF, PhD (IEEE Senior Member) is Professor at COMSATS University Islamabad, Wah Campus Pakistan. He has worked one year in Alpha Soft UK based software house in 1995. He is OCP in Developer Track. He is in teaching profession since 1996 to date. His research interests are Medical Imaging, Biometrics, Computer Vision, Machine Learning and Agriculture Plants. He is being awarded with COMSATS Research Productivity Award since 2011-2017. He served in TPC for IEEE FIT 2014-19 and currently serving as Associate Editor for IEEE Access, Guest Editor of Special Issues, and reviewer for well reputed journals.

RABIA SALEEM received the Ph.D. degree in pattern recognition from the COMSATS University Islamabad, Wah Campus. Since 2008, she has been in the education field. She is currently an Assistant Professor in Department of IT, Government College University Faisalabad. Her research interests include deep learning, machine learning, image processing, and natural language processing.

FADL DAHAN is an Associated Professor of Artificial Intelligence, Department of Management Information System, College of Business Administration - Hawtat Bani Tamim, Prince Sattam Bin Abdulaziz University, Al-Kharj, Saudi Arabia. Email: f.naji@psau.edu.sa. B.Sc: Thamar University, Yemen. M.Sc: King Saud University, Ph.D: Department of Computer Science, King Saud University. He is also a Faculty Member with the Department of Computer Science, Taiz University, Taiz, Yemen. His research interests include optimization and swarm intelligence.

ANUM MASOOD received the B.Sc. and M.Sc. degrees in Computer Science from the COMSATS University Islamabad, Islamabad, Pakistan, and the Ph.D. degree in Computer Science and Engineering from the Shanghai Jiao Tong University, Shanghai, China. She has worked as a Lecturer with the Department of Computer Science, COMSATS University Islamabad, Islamabad, Pakistan. She is a postdoctoral researcher with the Norwegian University of Science and Technology, Norway, and affiliated with PET Centre, St. Olav’s Hospital, Trondheim, Norway. In addition, she is a Research Fellow at the Radiology Department, Harvard Medical School, Harvard University, and Boston Children’s Hospital, Boston, U.S.A. She also worked as a visiting researcher with the Institute of Neuroscience and Medicine, Forschungszentrum Jülich; Institute for Cardiogenetics, University of Luebeck, Germany; and Liverpool John Moores University, Liverpool, U.K. Her research interests include medical image analysis, automated cancer detection, machine learning, and image processing.