

# A New Augmentation-based Method for Text Detection in Night and Day License Plate Images

<sup>1</sup>Pinaki Nath Chowdhury, <sup>2</sup>Palaiahnakote Shivakumara, <sup>1</sup>Umapada Pal, <sup>3</sup>Tong Lu and <sup>4</sup>Michael Blumenstein

<sup>1</sup>Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India.

Email:pinakinathc@gmail.com, umapada@isical.ac.in.

<sup>2</sup>Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. Email: shiva@um.edu.my

<sup>3</sup>National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China. Email: lutong@nju.edu.cn

<sup>4</sup>Faculty of Engineering and Information Technology, University of Technology Sydney, Australia. Email: Michael.Blumenstein@uts.edu.au.

## Abstract

Despite a number of methods that have been developed for License Plate Detection (LPD), most of these focus on day images for license plate detection. As a result, license plate detection in night images is still an elusive goal for researchers. This paper presents a new method for LPD based on augmentation and Gradient Vector Flow (GVF) in night and day images. The augmentation involves expanding windows for each pixel in R, G and B color spaces of the input image until the process finds dominant pixels in both night and day license plate images of the respective color spaces. We propose to fuse the dominant pixels in R, G and B color spaces to restore missing pixels. For the results of fusing night and day images, the proposed method explores Gradient Vector Flow (GVF) patterns to eliminate false dominant pixels, which results in candidate pixels. The proposed method explores further GVF arrow patterns to define a unique loop pattern that represents hole in the characters, which gives candidate components. Furthermore, the proposed approach uses a recognition concept to fix the bounding boxes, merging the bounding boxes and eliminating false positives, resulting in text/license plate detection in both night and day images. Experimental results on night images of our dataset and day images of standard license plate datasets, demonstrate that the proposed approach is robust compared to the state-of-the-art methods. To show the effectiveness of the proposed method, we also tested our approach on standard natural scene datasets, namely, ICDAR 2015, MSRA-TD-500, ICDAR 2017-MLT, Total-Text, CTW1500 and MS-COCO datasets, and their results are discussed.

**Keywords:** Augmentation, Gradient vector flow, RGB color space, Text detection, License plate detection, License plate recognition.

## 1. Introduction

Automatic driving, parking, toll point surveillance and security applications are the key areas that require special attention of researchers in the field of image processing and multimedia [1]. To make these applications successful, robust license plate detection is essential because license plate images captured for these applications pose many challenges, such as background variations, different heights or distances, low quality due to the dust, blur due to vehicle movements, illumination effect due to street lights etc. In addition, other challenges for license plate images captured during the night, such as poor quality due to low light and distortion i.e., from headlight illumination effects of many vehicles, make license plate detection much more complex [2]. When we compare night images with day images, one can expect that a day image has good quality in terms of high contrast, resolution and visibility, while a night one suffers from low contrast, low resolution and loss of visibility due to low light, head lights of vehicles and blur. License plate detection in night images is receiving special attention compared to day images because toll fee collection and monitoring illegal transportation automatically at night without human interventions is gaining importance. There are several sophisticated methods available in the literature [3] for license plate detection by addressing a few challenges mentioned-above, for example, the method in [1] is used for detecting dirty license plate images for high speed applications, the method in [4] is proposed for end-to-end car license plate detection and recognition with a deep neural network, which focuses on detecting accurate license plates by fixing exact bounding boxes, and the method in [2] deals with an efficient license plate recognition system using convolutional neural networks, which addresses the challenge of images with low lighting. It is noted from the existing methods that most of these do not consider both night and day images for experimentation. This is because the challenges posed by day images are different from those posed by night ones. In order to use independent methods, it is necessary to classify night and day images such that one can choose an appropriate approach to improve text detection performance. In this scenario, text detection performance depends on the success of a classification method. Thus the problem of handling both night and day images is complex and remains an unsolved issue for license plate detection.

There are methods for improving the quality of images which suffer from low resolution, contrast and low lighting. For example, Riesz's fractional based model [5] for enhancing license plate detection and recognition addresses the causes of illumination, weather conditions and night lighting [5]. The method in [6] ensembles ada boost cascades of 3L-LBPs classifiers for license plate detection in low-quality images and the method addresses the causes of low contrast, dusk, dirt, fog and distortions. In the case of enhancement-based methods, since enhancement is a preprocessing step for improving license plate detection performance, the success of license plate detection methods depends on the success of the enhancement step. To address this issue, methods have been developed for license plate detection without depending on a preprocessing step.

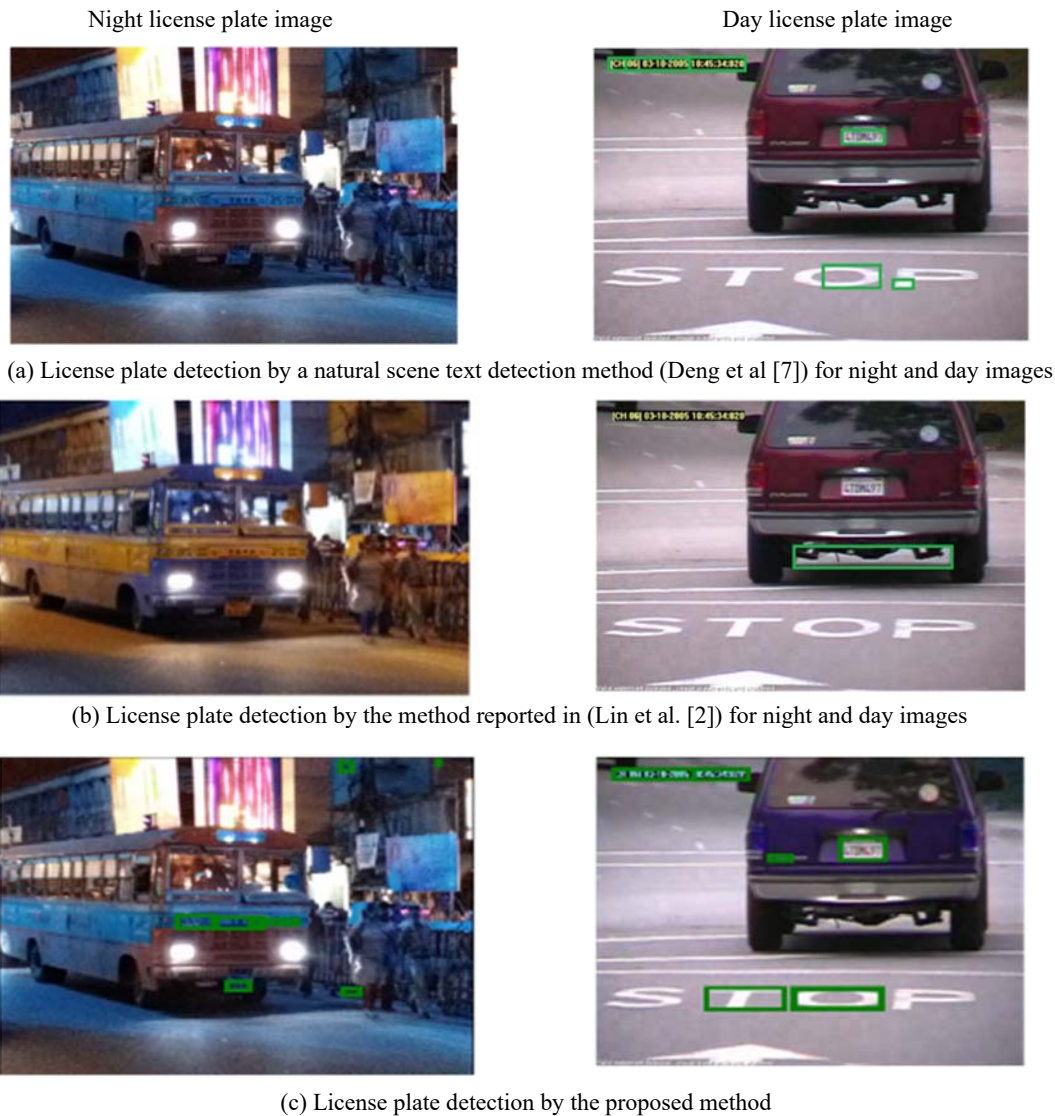


Fig. 1. Challenges for license plate detection in night and day images.

From the night and day sample images shown in Fig.1, it can be noted that night images are severely affected by head lights or street lights in addition to other challenges mentioned above, while day images do not have such effects and hence because of all these different effects, the problem is complex. It is evident from the illustration shown in Fig.1 that the method [7] which employs a powerful deep learning model for text detection in natural scene images does not detect license plates in night images, while it detects license plates in day images as shown in Fig.1(a). Note that according to natural scene text detection methods, license plate detection is a part of natural scene text detection [8] because the text exists naturally in both images. In the same way, the license plate detection method [2] which uses convolutional neural

networks does not detect license plates in night images, while it detects non-license plates as license plates in day images as shown in Fig.1(b). However, for both the night and day images featured in Fig.1, our proposed method detects license plates accurately as shown in Fig.1(c). In addition, it is observed from the results in Fig.1(c) that the proposed method detects not only license plates but also other texts in images. This is the advantage of the proposed method.

In this work, the scope is confined to license plate number detection in night and day images but not recognition. As discussed above, license plate detection in night and day images is complex and challenging. Furthermore, the success of recognition depends on the success of detection. Therefore, license plate detection is more important and is of interest to researchers.

## 2. Related Work

In this section, we review methods which detect texts in natural scene images, methods which detect license plates, and enhancement methods which improve poor quality images such that text detection performance improves. The existing methods are reviewed with respect to scope, objective and application of the proposed work to ascertain whether the existing methods can be used for night, day and night+day images, as reported in Table 1.

Table 1. Analysis of the existing methods on natural scene and license plate number detection with respect to night and day images.

Reference	Objective	Model	Night	Day	Night + Day
Panahi and I. Gholampour [1]	License plate	Connected component analysis	Yes	Yes	Yes
Lin et al. [2]	License plate	Deep learning	No	Yes	No
Li et al. [4]	License plate	Deep learning	No	Yes	No
Raghuandan et al. [5]	License plate	Fractional calculus	No	Yes	No
Shemarry et al. [6]	License plate	Local binary pattern	No	Yes	No
Deng et al. [7]	Natural scene	Deep Learning	No	Yes	No
Zamberletti et al. [8]	License plate and natural scene	Augmentation	No	Yes	No
Zhou et al. [9]	Natural scene	Deep learning	No	Yes	No
Shi et al. [10]	Natural scene	Deep learning	No	Yes	No
Ma et al. [11]	Natural scene	Deep learning	No	Yes	No
Zhang et al. [12]	Natural scene	MSER with Deep learning	No	Yes	No
He et al. [13]	Natural scene	Deep learning	No	Yes	No
Xue et al. [14]	Natural scene	Deep learning	No	Yes	No
Deng et al. [15]	Natural scene	Deep learning	No	Yes	No
Xu et al. [16]	Natural scene	Deep learning	No	Yes	No
Zhao et al. [17]	Natural scene	Deep learning	No	Yes	No
Liu et al. [18]	Natural scene	Deep learning	No	Yes	No
Bazazian et al. [19]	Natural scene	Deep learning	No	Yes	No
Khare et al. [20]	Video images	Motion vectors	No	Yes	No
Liu et al. [21]	Natural scene	Deep learning	No	Yes	No

Gupta et al. [22]	Natural scene	Deep learning	No	Yes	No
Yin et al. [23]	Natural scene	Clustering based model	No	Yes	No
Yao et al. [24]	Natural scene	Deep learning	No	Yes	No
Zhang et al. [25]	Natural scene	Deep learning	No	Yes	No
Tian et al. [26]	Natural scene	Deep learning	No	Yes	No
Koo and Kim [27]	Natural scene	Connected component analysis	No	Yes	No
Yin et al. [28]	Natural scene	MSER based model	No	Yes	No
Kang et al. [29]	Natural scene	Deep learning	No	Yes	No
Long et al. [30]	Natural scene	Deep learning	No	Yes	No
Liu et al. [31]	Natural scene	Deep learning	No	Yes	No
Liu et al. [32]	Natural scene	Deep learning	No	Yes	No
Xue et al. [33]	Natural scene	Gradient based model	Yes	Yes	Yes
Shivakumara et al. [34]	Natural scene and license plate	Fractional means	No	Yes	No
Asif et al. [35]	License plate	Edge based model	No	Yes	No
Xie et al. [36]	License plate	Deep learning	No	Yes	No
Yuan et al. [37]	License plate	Filter based model	No	Yes	No
Zhai et al. [38]	License plate	Deep learning	No	Yes	No
Ahmed et al. [39]	License plate	OCR based model	Yes	No	No
Zhang et al. [40]	Natural scene	Fusion	Yes	No	No
Chowdhury et al. [41]	License plate	Deep learning	Yes	No	No

It is noted from Table 1 that most methods focus either on text detection in natural scene images or license plate number detection but not on both. Zamberletti et al. [8] proposed augmented text characters and convolutional neural networks for text spotting from scene images as well as license plate images. In this work, the method spots words requested by a query through text detection in images. The main objective of the work is to investigate augmentation for character proposals to improve text detection performance without expecting a large number of samples and very deep architectures. However, it is observed from their experimental results that the method is inconsistent for natural scene and license plates images. Similarly, Shivakumara et al. [34] proposed a fractional means-based method for multi-oriented keyword spotting in video, natural scene and license plate images. However, the approach reports inconsistent results for images of different datasets. It is observed from the existing methods on license plate number detection mentioned in Table 1 that methods exist for detecting poor quality images caused by distortion. However, the methods do not consider poor quality images caused by low light or head lights of vehicles and blur.

It is noted from Table 1 that Panahi and I. Gholampour [1] proposed a method for license plate number recognition in day and night images based on adaptive clustering. It determines adaptive thresholds to handle day and night images. However, the threshold works well only for the simple images but not complex images because fixing a threshold is not easy for images affected by head lights and blur due to vehicle movements. Similarly, Xue et al. [33] proposed a method for text detection in blurred and non-blurred natural scene images based on gradient and filters. However, poor quality caused by low light at night is different from blur. In addition, in case of night images, we can expect uneven illumination effects

by head lights of vehicles along with blur information. Therefore, the method is inadequate for night images. Zhang et al. [40] and Chowdhury et al. [41] proposed enhancement approaches for improving the poor quality of night images such that text detection performance improves. However, the scope of the methods is limited to enhance the fine details in images but not text detection or license plate number detection. In addition, it is unclear how the methods perform for images with good quality and night images affected by the above-mentioned causes.

In summary, although there are a few methods for handling good and poor quality images, none of them consider both night and day images affected by uneven light effects in night and blur conditions. It motivated us to propose a method for addressing the challenges of both night/day license plate images. To the best of our knowledge, this is the first attempt to develop a unified method that combines hand-crafted features and deep learning models for achieving better results for text detection in night and day license plate images.

Thus, this work presents a new method which explores gradient, R, G and B color spaces, Gradient Vector Flow (GVF), and tackles the challenges posed by night/day license plate images. To reduce the impact of day and night effects, motivated by the method in [42] where it is shown that gradient operation enhances low contrast edges, we explore the gradient operation in different way in this work. This is because the gradient operation not only sharpens low contrast edges but also avoids the problem of foreground and background color changes. Inspired by the method in [8], which introduces an augmentation concept by expanding windows over character areas for text detection in day license plate and natural scene images, we explore the same concept with different operations in this work. The work presented in [40, 43] show that color information is important to handle the cause due to limited lighting conditions and hence we use R, G and B color spaces for collecting more information to restore missing pixels. It is noted that GVF generates unique patterns where there is text irrespective of poor and/or good quality images, we exploit the same GVF for obtaining the unique GVF arrow symmetry and loop pattern for separating texts from non-text information. In order to remove false positives and to improve performance of text detection in night/day license plate images, we use a learning-based recognition method [44], which removes false positives, and at the same time, the performance of text detection improves. The contributions of the proposed work are as follows.

The concepts proposed in [8, 40, 42, 43, 44] are well-known and used in many applications. It has motivated us to employ these methods for license plate detection in night and day images in this work. It is noted from [42] that GVF symmetry is defined for character segmentation in day images but not license plate number detection in day and night images. Similarly, color spaces are used for low light image enhancement in [40, 43] to improve text detection performance but not for both night and day images. In the same way,

MobileNet [44] was used for image segmentation and object detection but not license plate number detection in night and day images. We have modified the network such that it works well for license plate number detection in this work. Overall, the way the proposed method integrates the well-known concepts after modifications for finding a solution to these complex problems is new and different from the existing methods.

### 3. The Proposed Method

For each input image, the proposed method extracts the gradient information from R, G, B color spaces of input images. The reason to extract gradient information is that when the color of the foreground and background changes, the pixel values do not work. However, since the gradient operation works based on a transition from one color to another, it results in sharp edge pixels. In addition, since our focus is to extract stroke pixels (stroke with of characters in license plates) for differentiating text and non-text pixels, sharpening of the edges provide vital information for the classification of text and non-text pixels. However, due to the complexity of the images, this result includes non-text information at times. To reduce the effect of non-text edge information, the proposed method introduces an augmentation concept that performs a window expansion operation for each pixel over gradient images of the respective color spaces to extract stroke information, which in turn represents the edges of texts.

The reason to introduce augmentation is that it has the ability to extract meaningful information from images where information is hidden and is not evident [8]. In our case, pixels that represent character strokes in night images, which usually suffer from low contrast, are the hidden information. It is true that where there are edge pixels, there will be abrupt changes, which may be small, in the case of night images. To deal with this situation, there are two ways, namely, enhancing low contrast pixels, and collecting more information to extract meaningful material using neighboring pixels. In this work, we use an augmentation concept because when the enhancement step enhances low contrast pixels in day images, it reduces the gap between pixels of background and license plate numbers. This results in the misclassification of background pixels as those found in license plate numbers. However, in the case of the augmentation concept, unlike the enhancement step which alters pixel values, it grabs more information from the neighboring pixels by expanding the process. This results in meaningful information, which can distinguish character strokes from the background, irrespective of night and day images.



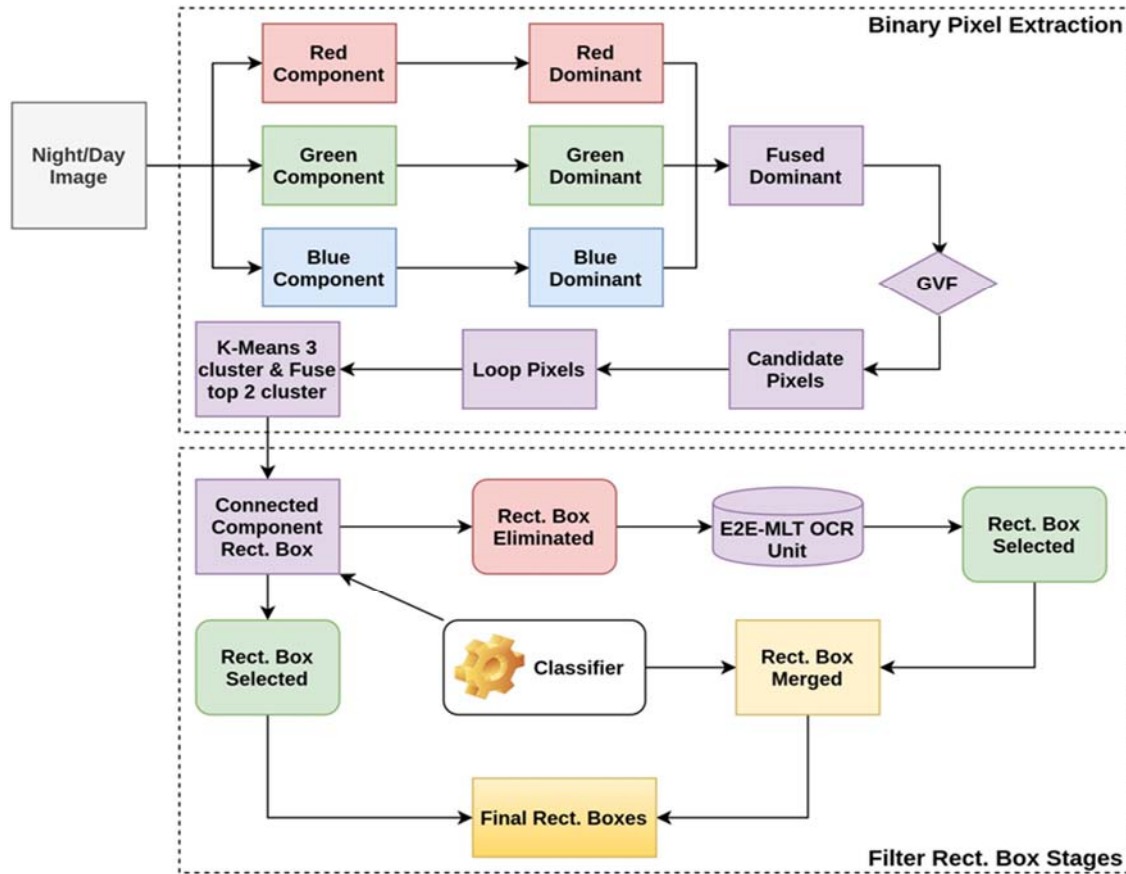


Fig. 2. The framework of the proposed method.

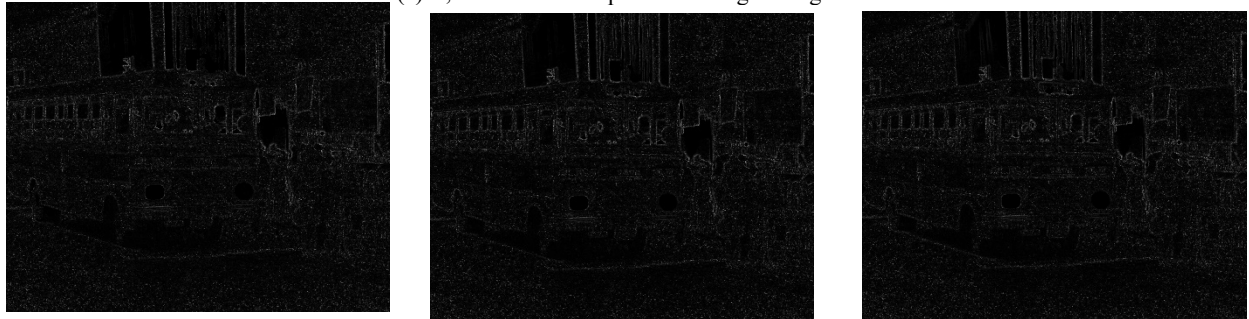
The augmentation step outputs dominant pixels for the input image. Since the problem is complex, dominant pixel detection alone is not adequate and hence it includes pixels that represent non-text information. It is noted that GVF provides the unique GVF pattern (GVF shows opposite arrow symmetry [42]) in the text portion. The proposed method extracts such an observation using GVF information for dominant pixels, resulting in candidate pixels. To improve the strength of the above GVF symmetry property for each candidate pixel's image, the proposed method checks one more unique GVF property called a loop pattern for representing the holes in the text. Sometimes, a hole can be found in non-text information. Therefore, to reduce pixels that represent non-text, for gradient values corresponding to holes of texts given by GVF loop patterns, the proposed method performs K-means clustering with  $K=3$  for classifying between text pixels and non-text ones. Since  $K=3$ , the cluster outputs three clusters containing high values, median values and low values, respectively, and then the proposed method performs a union operation for high and median clusters. This is because in general, low values represent non-texts, while high values represent text pixels. Hence, this process results in candidate text components. To remove false candidate text components, we explore deep learning-based recognition, which is a pre-defined architecture



[44]. The proposed method uses learning information for merging bounding boxes of candidate text components as well as removing false candidate text components, resulting in text detection in night/day license plate images. The pipeline of the proposed method can be seen in Fig.2.



(a) R, G and B color spaces for a night image



(b) Gradient images for the respective R, G and B color spaces in (a) of the night image



(c) R, G and B color spaces for the day image



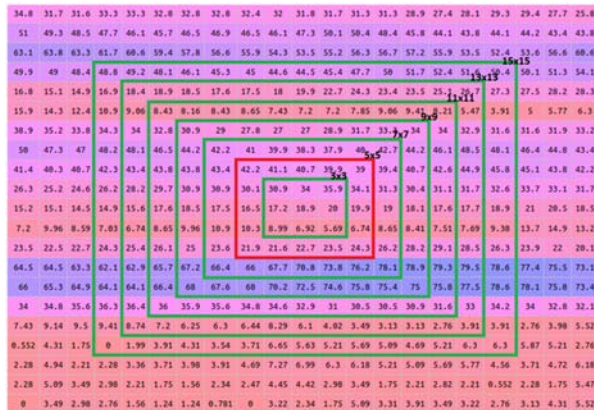
(d) Gradient images for the respective color spaces in (c) of the day image

Fig. 3. Studying the variations in pixel values in night and day license plate images.

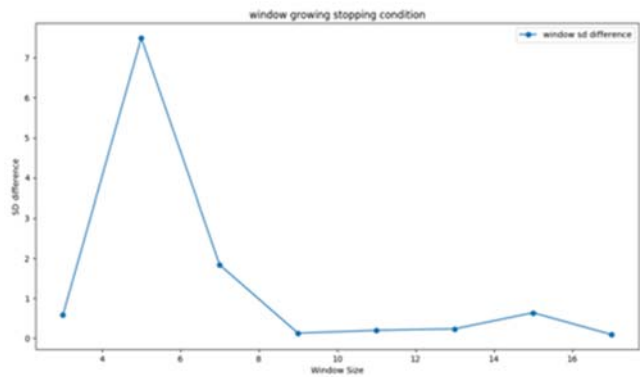
### 3.1. Augmentation for Dominant Pixel Detection

For each input image of night and day, the proposed method obtains color spaces, namely, R, G and B, and gradient images as shown in Fig.3 (a)-(b) and Fig.3(c)-(d), respectively. It is observed from Fig.3 (a)-Fig.3(d) that although the images are affected by night and day, color and gradient information provide fine details of the edges. It is also true that the thickness of character strokes in a license plate remain constant [42]. Similarly, it is true for texts in natural scene images [45]. In order to exploit this observation, inspired by the method in [8] where window expansion is used for increasing the number of patches to extract significant character information, we propose to use a window expansion method which is called Augmentation to extract dominant pixel information from images.

For each pixel in the input image, the proposed method fixes a square window of  $3 \times 3$ , and calculates the standard deviation, which denotes iteration-1. This process repeats for the window of dimensions  $5 \times 5$ , which denotes iteration-2. The proposed method finds the absolute difference between the standard deviations of iteration-1 and iteration-2. It is followed by the difference between iteration-2 and iteration-3 as illustrated in Fig.4 (a), where one can see the process of window expansion. This window expansion process terminates when the difference starts decreasing as shown in Fig.4 (b), where it is noted that the window expansion process terminates at a  $5 \times 5$  window for the night image. The proposed method calculates the distance from the starting pixels to the terminating pixel. This results in a distance matrix for all the pixels in the input image.



(a) Window expansion for dominant pixel selection

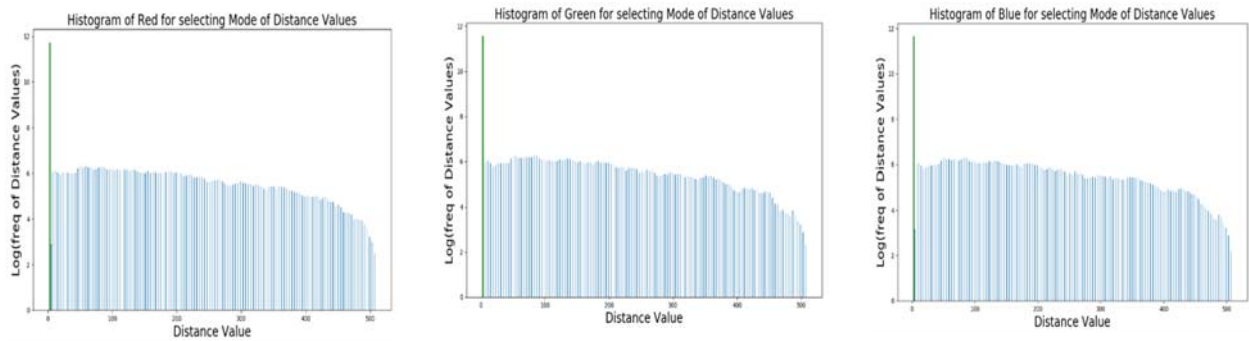


(b) Stopping criteria for window expansion

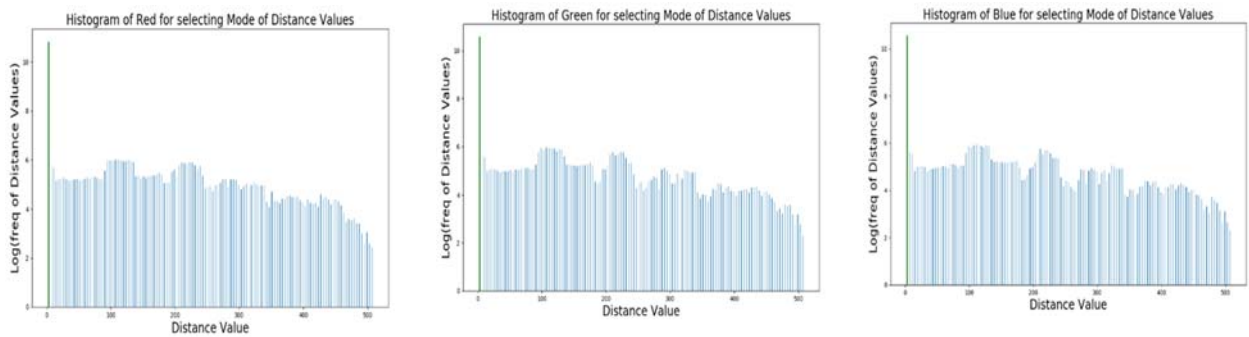
Fig. 4. Augmentation for dominant pixel detection of a night image

As mentioned above, the distance matrix contains constant values, which represent character strokes in license plate images. For extracting pixels that represent character strokes, the proposed method performs a histogram operation to find the pixels which contribute to the highest peak as shown in Fig.5 (a) and Fig.5 (b), respectively, for R, G and B color spaces. The pixels that represent the highest peak are displayed as white for the purpose of visualization in Fig.6 (a) and Fig.6 (b) for R, G and B color spaces, respectively,

which results in dominant pixels. In Fig.6 (a) and Fig.6 (b), it is noted that the result includes pixels of character strokes as well as pixels of strokes of background objects (for example, cars). This is expected as the thickness of background objects sometimes overlap with the thickness of character strokes. Fig.6 (a) and Fig.6 (b) show that the day image misses a few pixels of characters and includes a few stroke pixels of non-text information, while the night image misses a few pixels of characters and includes more non-text information compared to the day image. This is because in the case of night images, pixel values are very close to each other despite there being a difference between edge and background pixels. Thus it is not so easy to separate edge pixels from the background. Therefore, many background pixels in night images satisfy the condition used in the augmentation, and hence more background pixels are classified as dominant ones for night images compared to day images. As a result, there is a need of a step which can eliminate non-text pixels as well as restoring missing text pixels in the input image, which will be discussed in the subsequent section.

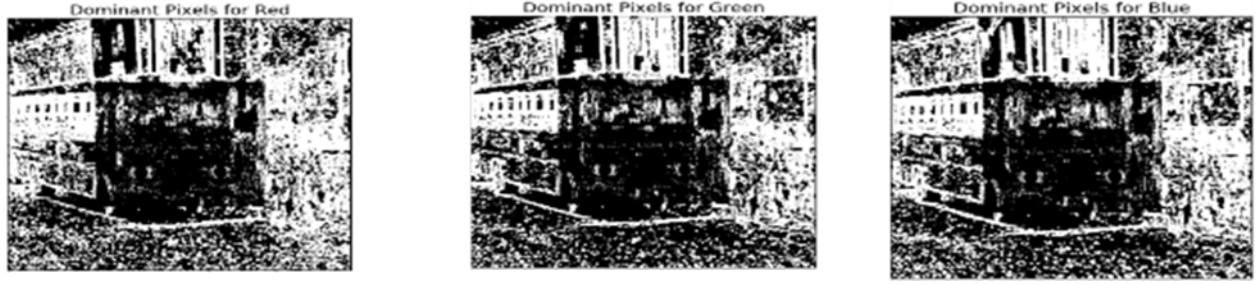


(a) Histogram for the distance values of the respective R, G ad B color spaces of the night image



(b) Histogram for the distance values of the respective R, G ad B color spaces of the day image

Fig. 5. Choosing the pixel values which contribute to the dominant peak in the histogram for night and day images



(a) Dominant pixels for the respective R, G and B color spaces of a night image



(b) Dominant pixels for the respective R, G and B color spaces of a day image

Fig. 6. Dominant pixel detection of the proposed method for night and day images.

### 3.2. Gradient Vector Flow Symmetry for Candidate Pixel Detection

As noted from Fig.6 (a) and Fig.6 (b), the missing pixels in license plate images can be restored by fusing the dominant pixels in R, G and B color spaces. This is because sometimes, R may contain pixels, which may disappear either in G and B color space or vice versa. Therefore, we propose a fusion operation as defined in Equation (1) and Equation (2). The fusion operation checks whether the window of size  $3 \times 3$  over R, G, B color spaces contain at least one dominant pixel or not. If it is present, the fusion operation performs a union to combine the dominant pixels in R, G and B color spaces, as shown in Fig.7 (a) and Fig.7 (b) for night and day images, respectively. In those Figures we can see that a few missing pixels from Fig.6 (a) and Fig.6 (b) are restored. However, it is noted from Fig.7 (a) and Fig.7 (b) that this step includes more non-text pixels in the night and day images, at the same time, the pixels of license plates are restored. Due to the night image effect, the fusion operation alone may not be sufficient to restore the missing dominant pixels for all situations. To overcome this problem, motivated by the statement mentioned in [42], which says if there are edge pixels in the image, one can expect opposite GVF arrows.

Therefore, we propose GVF opposite arrow symmetry criteria to capture more information. To achieve this, for each dominant pixel, the proposed method checks the GVF opposite arrow symmetry as illustrated in Fig.7(c) for all the 8 neighbors of its pixels in the GVF image of the input image. If a neighboring pixel satisfies the GVF opposite arrow symmetry, the proposed approach considers it as a dominant one, which is called candidate pixel detection. The effect of this step for night and day images is shown in Fig.7 (c),

where it can be seen that most of the license plate pixels are restored. However, this process includes non-text pixels also.

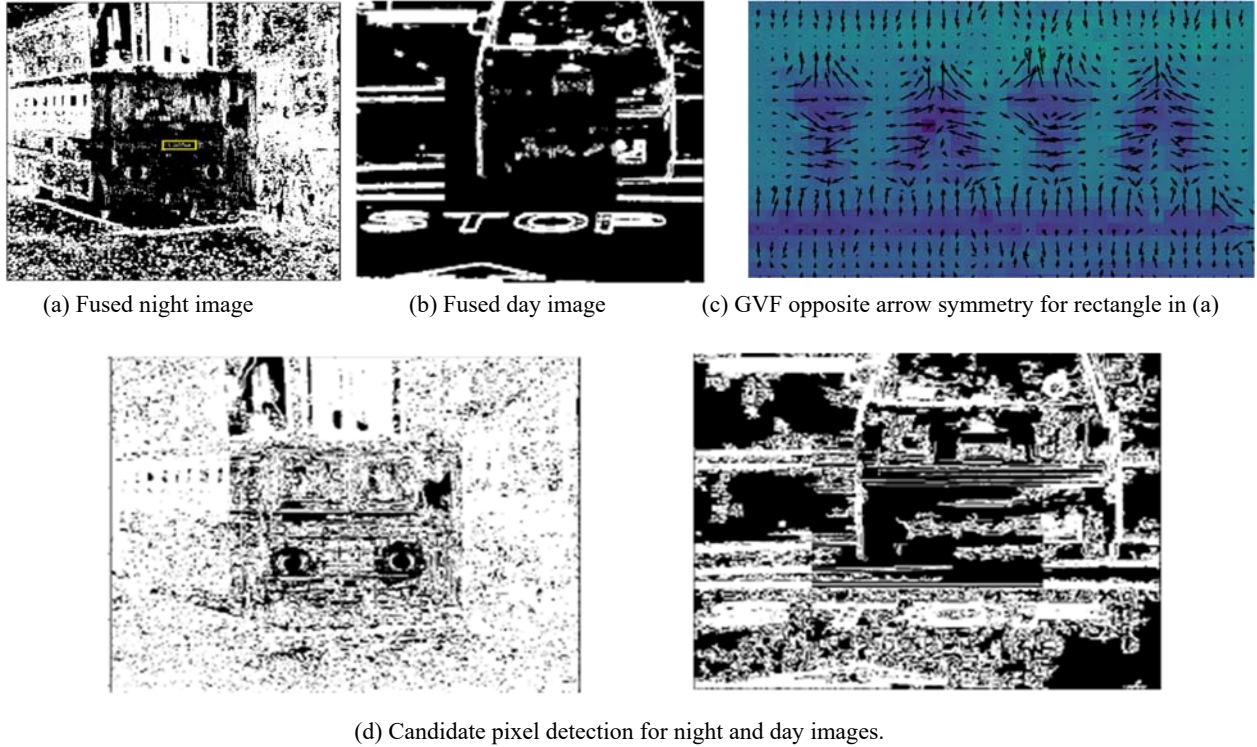


Fig. 7. The steps for candidate pixel detection of night and day license plate images.

$$\text{If } \|X_{(x,y)}^{Red}\|_1 \geq 1 \text{ and } \|X_{(x,y)}^{Green}\|_1 \geq 1 \text{ and } \|X_{(x,y)}^{Blue}\|_1 \geq 1 \quad (1)$$

$$X_{(x,y)} = X_{(x,y)}^{Red} \vee X_{(x,y)}^{Green} \vee X_{(x,y)}^{Blue} \quad (2)$$

where,  $(x, y)$  denotes pixel location,  $\{X_{(x,y)}^{Red}, X_{(x,y)}^{Green}, X_{(x,y)}^{Blue}\} \in R^{3 \times 3}$  represents dominant pixels of R, G and B color spaces, respectively, and R denotes pixel values of a window of size  $3 \times 3$ .

### 3.3. GVF Loop Pattern for Candidate Components Detection

It is true that if there are characters in images, some cursive shapes may be expected, which create a hole. Since the above step includes non-text pixels along with license plate pixels, we propose to extract the above observations for each candidate pixel in the image to remove non-text candidate pixels. As discussed in the previous section, if there is a hole or cursive shape in the image, the GVF arrow generates a pattern-like loop as shown in Fig.8 (a), where one can see GVF arrow loop patterns in the areas of a text region. The proposed approach considers GVF arrow direction and standard deviation of 8 neighbors of centered candidate pixels to detect pixels that represent GVF arrow loop patterns as the algorithmic steps defined in Algorithm-1. For each candidate pixel that contributes to the loop pattern, our method extracts gradient



values corresponding to pixels of the loop pattern. It is true that the pixels which represent the loop pattern have high gradient values, while non-text pixels have low gradient values. However, due to night image effects, one cannot expect high values for the pixels of the loop pattern at all times. Therefore, we propose to use K-means clustering with  $K=3$  for classifying the pixels which represent license plates in images. The output of K-means clustering on the gradient values of pixels of the loop pattern results in three clusters as shown in Fig.8 (b) and Fig.8 (c) for night and day images, respectively. Since K-means clustering is unsupervised, the proposed method defines the cluster which gives a high mean as the Max cluster, low mean as a Min cluster and the other one as a Median cluster. To tackle the challenge of night images, the proposed method combines the pixels in the Max and Median clusters with a union operation, as shown in Fig.8 (b) and Fig.8 (c) for night and day images, respectively. Since the Min cluster contains the pixels of non-text, our method discards those pixels as shown in Fig.8 (b) and Fig.8 (c), where almost all non-text pixels are classified into the Min cluster. The result of the union operation for Max and Median clusters is considered as candidate component detection because the step helps us to group text pixels into components as shown in the result of the union operation in Fig.8 (b) and Fig.8 (c) for night and day images, respectively. The algorithmic steps for finding pixels of the loop pattern and candidate components are defined in Algorithm-1.

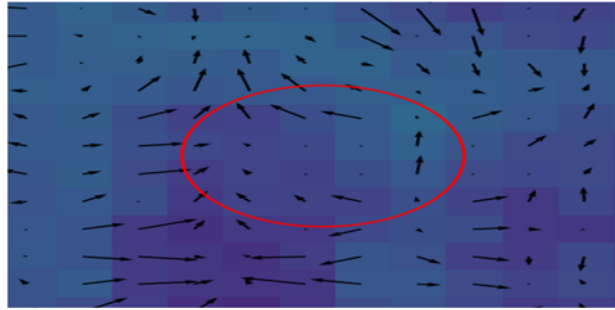
#### Algorithm-1: Candidate Component Detection

$L$  is an array which contains the pixels of a loop pattern,  $p_{(x,y)}$  denotes the candidate pixel in the image,  $C$  contains candidate pixels given by the step presented in the previous section,  $M_{(x,y)} \in \mathbb{R}^{3 \times 3}$  represents a square matrix centered at  $(x,y)$  with dimensions  $(3 \times 3)$ . Note that  $M_{(x,y)}$  contains gradient values corresponding to the pixels of loop patterns,  $\sigma_{M_{(x,y)}}$  denotes the standard deviation of the  $M_{(x,y)}$ , and  $T$  is a temporary list which contains temporary locations while finding loop pixels.

```

1.  $L \leftarrow []$  ▶ create an empty list
2. For each  $p_{(x,y)}$  in  $C$  do
    a.  $T \leftarrow []$  ▶ create temporary empty list
    b. append  $(x,y)$  to  $T$ 
    c. while (True) do
        i.  $(x',y') \leftarrow$  use GVF direction at  $(x,y)$  to move to the next point
        ii. if  $(x',y') == (x,y)$  then
            Break
    d. End if
    e. if  $\sigma_{M_{(x,y)}} \geq$  threshold then
        i. Break
    f. End if
        i. if  $(x',y')$  in  $T$  then
            For  $(i,j)$  in  $T$  do ▶ append all points in  $T$  to  $L$ 
                append  $(i,j)$  to  $L$ 
            1. End for
            2. Break
    g. End if
    h.  $(x,y) \leftarrow (x',y')$ 
3. End while
4. End for

```



(a) GVF Loop Pattern extraction for the holes of the characters in the night license plate image.



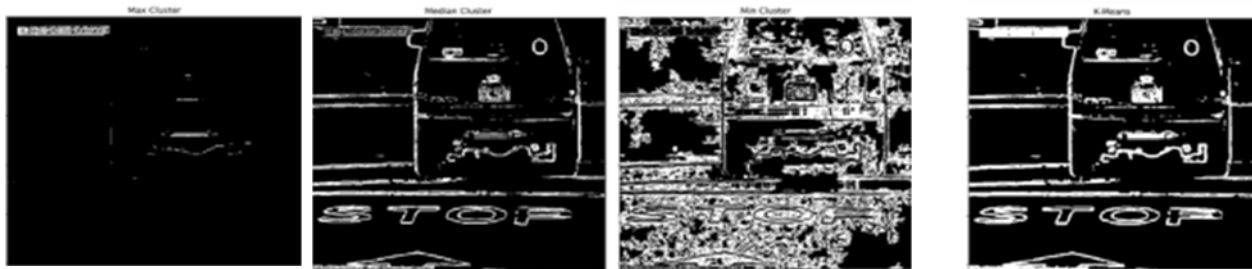
Max cluster

Median cluster

Min cluster

Union of Max and Median cluster

(b) Candidate component detection for the night license plate image



Max cluster

Median cluster

Min cluster

Union of Max and Median cluster

(c) Candidate component detection for the day license plate image

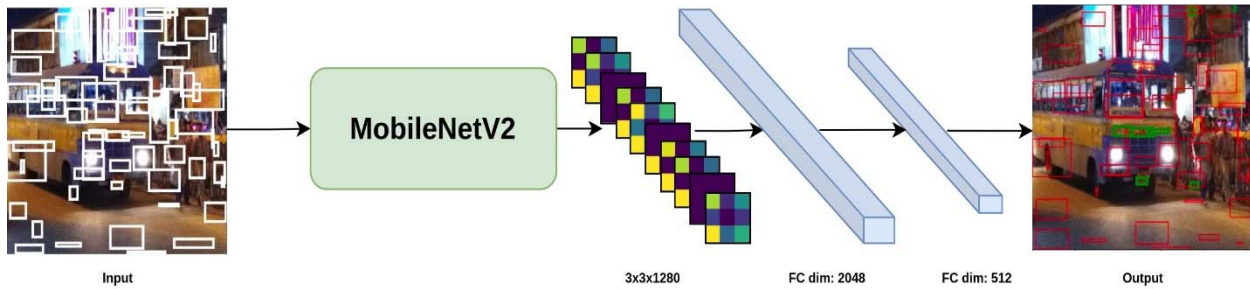
Fig. 8. The steps for candidate component detection

### 3.4. Text Detection in Night and Day License Plate Images

The step presented in the previous section outputs candidate components, which include the whole license plate of the image. Therefore, in this step, the proposed method considers each candidate component as the input for detecting texts by eliminating false positives. Since the whole license plate is one component, we explore the use of classification and recognition through a deep learning model for detecting texts while removing non-texts. The architecture for learning is shown in Fig.9, where the architecture comprises a MobileNetV2 deep learning model [44], which outputs binary results for each input component as shown in Fig.10 (a), where we can see each box is considered as one candidate input component. The deep learning model uses MobileNetV2 for feature extraction and a custom fully connected classifier, which outputs a binary decision, where 1 represents absolute certainty of the input image containing text and vice-versa.



The MobileNetV2 is initialized using ImageNet weights, and the custom fully connected classifier is initialized using Xavier's initialization method.



**Fig. 9. Architecture for text detection in night and day license plate images.** White boxes in the images are input and the boxes with green and red colors are the output. The green color indicates a correct license plate number and the red color indicates non-text.

For creating training data for this deep learning approach, we use the ICDAR 2017 MLT training set [13] because it provides labeled text regions. The labeled information is used to generate crop boxes containing license plates or text information, and a random region of backgrounds along with it. As a result, the cropped images created with text and background information become positive samples for our training data. In addition, the proposed method crops random background regions from ICDAR 2017 MLT, where the IOU of cropped images that are equal to 0 become negative samples in our training data. The effect of recognition through the deep learning model is illustrated in Fig.10 (a) and Fig.10 (b) for night and day images, respectively. It is observed from Fig.10 (a) and Fig.10 (b) that the model removes false positives, merging boxes of text components as one component and the final text detection for the night and day images. It is noted from Fig.10 (a) and Fig.10 (b) that the proposed method detects not only license plates, but also other texts in images. This shows that the proposed method is independent of text and hence it works well for text detection in natural scene images. The algorithmic steps for text detection using recognition are presented in Algorithm-2.

**Algorithm-2: Deep learning model for Text detection**

B contains a list of initial rectangular boxes, CB denotes Correct Box, which contains a list of correct boxes and it is the final output. WB denotes wrong box, which are classified as false positives by the proposed method. NPB denotes new potential boxes, which contains a list of boxes initially rejected by the classifier, but the text recognizer found some text inside them.  $Area(b_i, b_j)$  denotes an area of the resulting rectangular box when we merge the area of two adjacent rectangular boxes, say,  $b_i$  and  $b_j$ .  $Height_{merged\_box_{ij}}$  indicates the height of the resulting rectangular boxes' box (i.e. merging  $b_i$  and  $b_j$ ),  $Width_{merged\_box_{ij}}$  indicates the width of the resulting rectangular boxes' box (i.e. merging  $b_i$  and  $b_j$ ).  $\min(a, b)$  indicates the minimum area of the boxes.

1.  $CB \leftarrow [ ]$ ;  $WB \leftarrow [ ]$ ;  $NPB \leftarrow [ ]$       ► initialize empty list
2.  $Classifier() \leftarrow$  deep learning classifier to predict whether a box is text / non-text
3. For each  $b$  in B do
  - a. if  $Classifier(b)$  predicts text then
    - i. append  $b$  to CB
  - b. else

- i. append b to WB
  - c. End if
4. End for
5. For each b in WB do
  - a. if E2E-MLT\_OCR predicts any character inside b then
    - i. append b to NPB
  - b. End if
6. For every pair  $(b_i, b_j)$  in NPB do
  - a.  $b_{merged} \leftarrow \text{merge}(b_i, b_j)$
  - b. if  $\text{Height}_{merged\_box_{ij}} < \text{threshold}$  and  $\text{Width}_{merged\_box_{ij}} < \text{threshold}$  then
    - i. replace  $(b_i, b_j)$  in NPB with  $b_{merged}$
  - c. End if
7. End for
8. For each b in NPB do
  - a. if Classifier(b) predicts text then
    - i. append b to CB
  - b. End if
9. End for



Candidate components



False components removal



Merging the components



License plate detection

(a) The process of eliminating false candidate components for text detection in a night license plate image.



Candidate components



False components removal



Merging the components



License plate detection

(b) The process of eliminating false candidate components for text detection in a day license plate image.

Fig.10. Text detection in night and day license plate images

## 4. Experimental Results

Since the aim of the proposed work is to detect texts in night, day license plate images as well natural scene images, we consider license plate image datasets and the standard benchmark datasets of natural scene images in this work. However, since there is no standard dataset for night images, we created our own dataset for text detection, which includes images of different timings from evening to night, different illumination effects, different orientated text images, images affected by perspective distortions, multiple vehicles in single images, etc., as shown in sample images from Fig.11. To test the robustness of the

proposed method, we consider the standard datasets of day license plate images, namely, UCSD [8, 34]. The dataset includes images with blur, low resolution, noise and images with good focus as shown in the sample images in Fig.11.

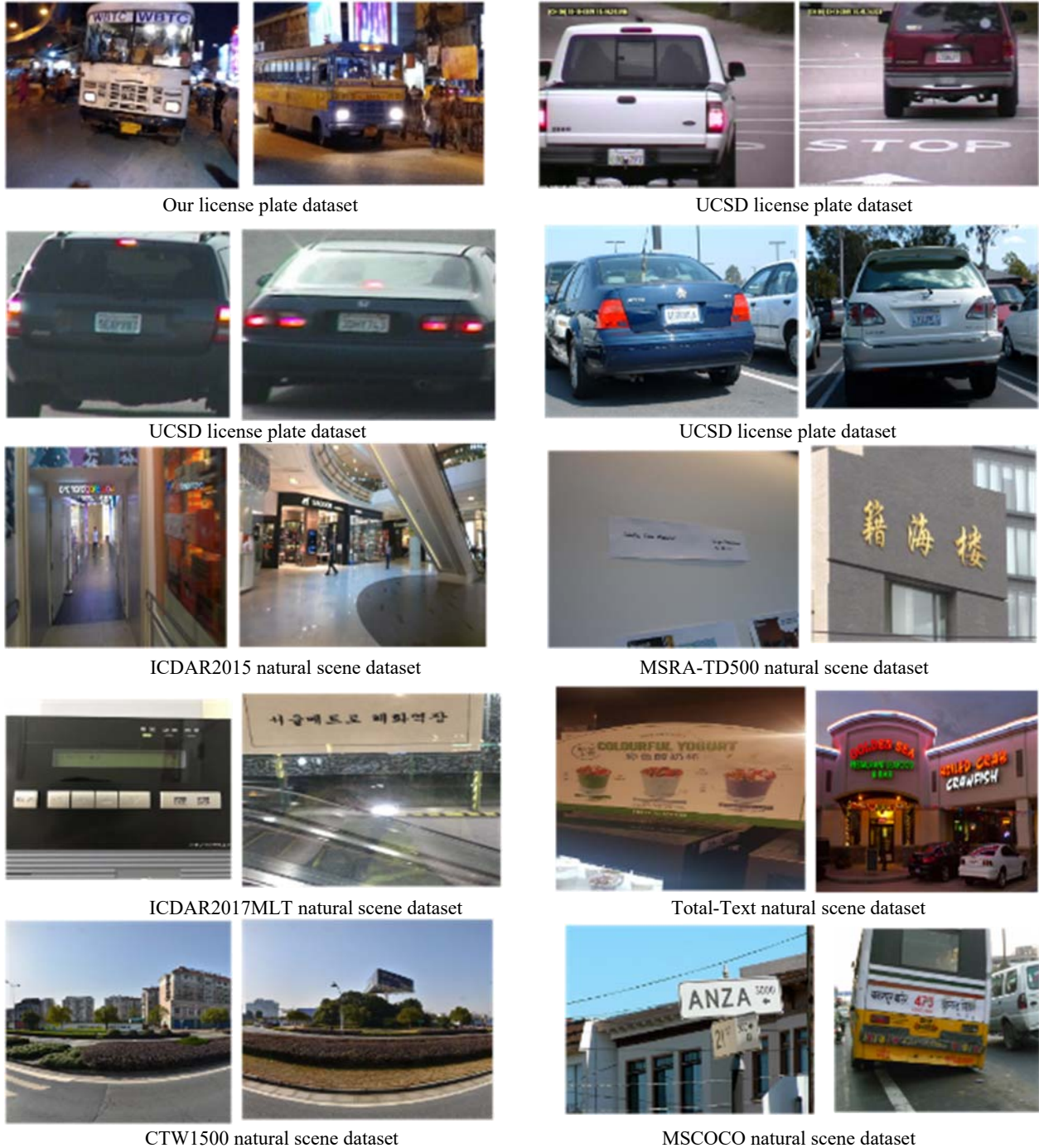


Fig.11. Sample images of the different license plate and natural scene databases.

As discussed in the Proposed Methodology Section, our method is independent of text and image, we thus consider the benchmarks of natural scene text images for evaluation in this work, namely, ICDAR 2015 [46] which contains horizontal texts of different variations, MSRA-TD500 [13] which contains arbitrary oriented texts, ICDAR 2017-MLT [13] which contains texts of multiple languages, Total-Text [47] which provides curved texts of different scripts, CTW1500 [32] which contains curved texts, and MSCOCO [48] which contains curved texts with very complex backgrounds. The sample images for the above-mentioned datasets are shown in Fig.11, where one can understand the nature of the datasets. The statistical details including characteristics and numbers of testing samples for all the datasets are listed in Table 2, where different attributes of datasets can be seen. Table 2 shows that the datasets considered for experimentation and evaluation have large variations in terms of resolution and artifacts such as blur, perspective distortion, orientation, script and illumination. Therefore, we believe that these datasets provide the possible variations for evaluating the robustness, effectiveness, usefulness of the proposed method.

Table 2. Statistical details of the different databases including our license plate dataset and natural scene datasets.

Databases	Testing samples	Resolution	Blur	Script MLT	Imbalanced Illumination	Orientation	Perspective
License Plate- Our Dataset-Night	200	Min: 2067×2066 Max: 3120×4160	✓	✓	✓	✓	✓
License Plate-UCSD-Day [8]	1695	Min: 241×241 Max: 463×685	✓	✗	✗	✗	✗
Natural Scene- ICDAR 2015-Day [46]	500	Min: 720×1280 Max: 720×1280	✗	✗	✗	✓	✓
Natural Scene- MSRATD-500-Day [13]	300	Min: 864×1296 Max: 1296×1728	✗	✓	✗	✓	✓
Natural Scene-ICDAR2017MLT-Day [13]	9000	Min: 302×200 Max: 4032×3024	✗	✓	✗	✓	✓
Natural Scene-Total Text-Day [47]	300	Min: 165×240 Max: 3456×5184	✗	✓	✗	✓	✓
Natural Scene- CTW1500-Day [32]	500	Min: 2048×2048 Max: 2048×2048	✗	✓	✗	✓	✓
Natural Scene-MSCOCO-Day [48]	900	Min: 265×444 Max: 640×640	✗	✓	✗	✓	✓

To measure the performance of the proposed method, we follow the standard metrics and evaluation schemes as in [13], and the measures are defined in Equation (3). We define the measures as follows. If a rectangular box has text inside a closed bounding box, then those boxes are considered to be True positives (TP). Similarly, if there is a Rectangular Box without any text inside it, we consider this as a False Positive (FP). If there is text but no rectangular boxes, then such text is missed by the proposed method and labeled as a True Negative (TN). Based on this, the different measures are used as follows.

$$Precision = \frac{TP}{TP+FP}; Recall = \frac{TP}{TP+TN}; F\_Score = \frac{2*Precision*Recall}{(Precision+Recall)} \quad (3)$$

To show the effectiveness of the proposed method, we choose the existing methods of natural scene text detection, the methods of license plate detection, and the methods for low light images (for improving text



detection performance), to compare them with the proposed method. The methods, namely, EAST [9], SegLink [10] and Pixel\_Link [7], are proposed for text detection in natural scene images by exploring different deep learning models. Note that the method in [2] is proposed for license plate detection using convolutional neural networks. The methods, namely, Zhang et al. [40] and Chowdhury et al. [41] proposed approaches to enhance fine details in low light and night images to improve text detection performance for poor quality images. The method in [40] fuses frequency and spatial information of images, while the method [41] uses a deep learning model for enhancement. The reason to choose the above-stated methods is that they [7, 9, 10] are popular and they are also the state-of-the-art methods for text detection in natural scene images. In addition, the methods address several challenges of scene text detection as license plate detection. The method in [2] is developed for addressing challenges of license plate detection in low light images as in the proposed method. Since the objectives of the proposed method and the two existing enhancement methods [40, 41] are the same, we implemented these two existing methods to undertake a comparative study with our method.

In summary, [2, 7, 9, 10] can be categorized as the methods for day images, and [40, 41] can be considered as the methods for night images for experimentation. It is expected that the approaches [40, 41] should achieve better results for night images, while the methods [2, 7, 9, 10] are expected to achieve better results for day images. For the datasets of night and day images together, all the existing methods should report neither high nor low compared to the respective individual datasets.

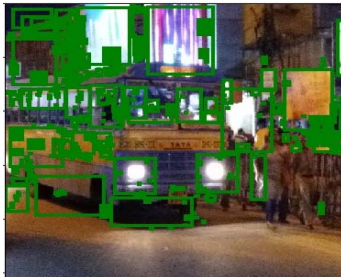
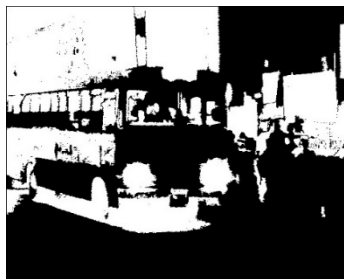
#### **4.1. Evaluating the Key Steps of the Proposed Method**

In this work, the use of color spaces for collecting vital information, introducing an augmentation concept for detecting dominant pixels, and exploring deep learning models for merging boxes of text components to detect texts in night/day license plates and natural scene images are the key steps. In order to analyze the contribution of each key step, we conducted experiments for each step on our night dataset to calculate the measures as reported in Table 2. For the experiments on color spaces, the proposed method considers individual color space as the input for detecting text in the images. In this case, the proposed method misses the fusion step discussed in Section 3.2 for detecting candidate pixels. The proposed method considers dominant pixels given by individual color spaces as candidate pixels for candidate component detection. When we look at the results of individual color spaces reported in Table 3, the variation in measures is nominal. However, the score of individual color spaces is lower than that of the proposed method. This is due to some of the vital information of fusion being missed. This shows that the color spaces contribute equally for achieving the best score by the proposed method. In the case of the augmentation step, the proposed method performs a histogram operation on gradient values of color spaces rather than distance values for detecting dominant pixels in images. Since this step misses stroke information, one can expect

more false positives and hence the precision is low compared to that of the proposed method without the augmentation step. However, the step gives more information, and the recall of the proposed method is almost the same as reported in Table 3. It is true that precision is more important than recall in terms of reliability and trust for real-world applications. This shows that the step of augmentation contributes to the reliability of the proposed method. In the case of merging boxes of text components to get full text lines or words, the proposed method considers all candidate components as actual text information after removing false positives. In the absence of merging, since text boxes do not contain sufficient text information, the proposed recognition step does not detect text boxes most of the time. As a result, the recall is very low compared to that of the proposed method as reported in Table 3. However, the precision is high compared to the proposed method because the step without merging retains only the boxes that contain correct texts and reduces the number of false positives. Therefore, the merging step is important to improve the recall and the overall F-score. From the above discussions, it can be concluded that the key steps mentioned above contribute significantly to achieve the best results by the proposed method.

The results of MSER Candidate components

License plate detection



(a) Proposed method with MSER for the night image



(b) Proposed method with MSER for the day image

Fig. 12. Performance of the proposed method by considering dominant points detection by MSER as input.

Similarly, to test the effectiveness of the dominant pixel step of the proposed method for night and day images, we replaced the dominant pixels by the output of the Maximally Stable Extremal Regions (MSER) approach, which was used in [23] for candidate pixels detection in natural scene images, for calculating measures on our datasets. Sample qualitative results of the Proposed Method with MSER on night and day images are shown in Fig. 12 (a) and Fig. 12 (b), respectively, where it is noted that the step detects license plate numbers for day images, whilst it misses those in night images. This could be expected as MSER is

appropriate for images, which are good quality and high contrast, but not for poor quality and night images. It can be verified by the quantitative results of the proposed method with MSER reported in Table 3, where the F-score is lower than that of the proposed method because the recall is too low. This shows that MSER misses text in night images. Therefore, the proposed method with dominant pixels given by augmentation is better than the conventional MSER.

Table 3. Assessing the contribution of key steps of the proposed method

Key steps	Precision	Recall	F-Score
Proposed method	0.81	0.72	0.76
Proposed method with only red color space	0.78	0.71	0.74
Proposed method with only green color space	0.81	0.71	0.75
Proposed method with only blue color space	0.79	0.73	0.75
Proposed method without augmentation step	0.78	0.71	0.74
Proposed method without boxes merging step	0.87	0.56	0.68
Proposed Method with MSER [23]	0.87	0.63	0.73

#### 4.2. Experiments on Night and Day License Plate Datasets

Qualitative results of the proposed method for our night license plate dataset and the benchmark day license plate datasets are shown in Fig.13, where it is noted that the proposed method detects license plates well for images affected by different scenarios, such as night effect, blur, blur + noise and well-focused images. This shows that the proposed method is capable of handling different challenges especially night and day license plate images. The reason for this success is that the combination of color spaces + gradient information, augmentation for dominant pixel detection, GVF-based patterns for candidate pixel/component detection, and the deep learning-based classifier discussed in the above can withstand such challenges.

For evaluating the proposed and existing methods in terms of quantitative results, we conduct experiments on our Night, UCSD day datasets and Night with UCSD together (Night + UCSD), which results in three sets of data for experimentation. The results of the proposed and existing methods on the three datasets are reported in Table 4, where it is observed that for night datasets, all the methods including the proposed method, report poor results in terms of Recall and F-measure compared to Day and Night + Day datasets. As expected, most methods including the proposed method score well for day images, low for night images and neither low nor high for Night + Day images. This shows that the night datasets are challenging compared to day datasets for achieving better results. When we compare the performance of the methods of day images [2, 7, 9, 10] and night images [40, 41], the methods of night and day images achieve the best results for the respective datasets. For instance, the method [40] is the best at Precision for night images while the method [9] is best at Precision for day images. But for the datasets of night + day images, the existing methods score poorly compared to the respective individual datasets. Therefore, one can conclude that the existing methods are inadequate to handle both night and day images. On the other hand, the



proposed method is consistent for night, day and Night + Day images especially on Recall and F-measure. However, the precision is lower than the existing methods for the datasets. This is mainly because sometimes, the proposed method gets confused with the parts of cars in images, and hence they are classified as actual license plates. Thus, the precision is low. We believe detecting license plates in different conditions and obtaining stable results are important for the applications of license plate detection. Therefore, we can infer that the proposed method is useful and capable of handling challenges of both night and day images.

To take advantage of the method of night and day images, which performs well on the respective datasets, we need the classification of night and day images such that one can choose an appropriate method to achieve better results. However, in this case, the performance of the methods depends on the success of classification. It is noted that classification of night and day images is not easy because sometimes both images can share the effect of the same effects, such as low resolution, low contrast and complex background. In addition, a single image can contain multiple portions which suffer from different qualities. Therefore, developing a method that can work well for night and day is reliable and stable.

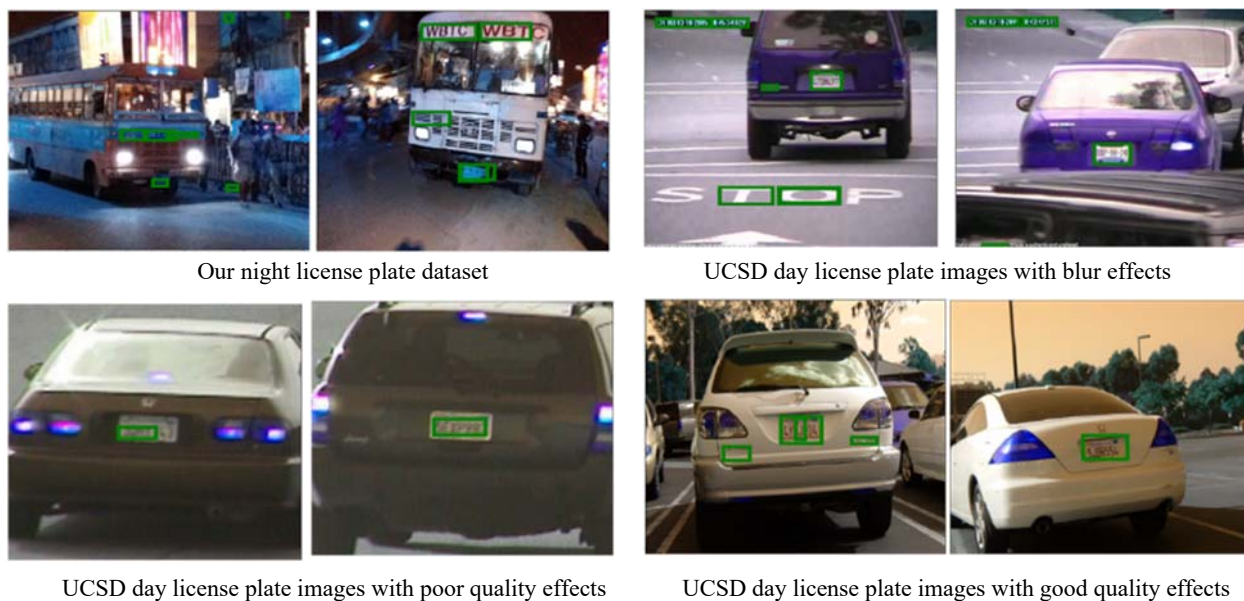


Fig.13. Qualitative results of the proposed method for different license plate datasets.

Table 4. The performance of the proposed method on different night and day license plate datasets

Methods	Our Night dataset			UCSD-Day dataset			Night + Day datasets		
	P	R	F	P	R	F	P	R	F
EAST[9]-Day	0.88	0.37	0.52	<b>0.99</b>	0.51	0.67	0.93	0.44	0.60
SegLink[10]-Day	0.93	0.36	0.49	0.93	0.61	0.74	0.93	0.48	0.63
PixelLink[7]-Day	0.96	0.54	0.69	0.98	0.65	0.78	<b>0.97</b>	0.59	0.73
LP[2]-Day	0.72	0.33	0.46	0.66	0.35	0.46	0.69	0.34	0.45
Fusion [40]-Night	<b>0.97</b>	0.60	0.74	0.94	0.70	0.80	0.96	0.65	0.77

U-Net [41]-Night	0.82	0.41	0.55	0.92	0.69	0.74	0.87	0.52	0.65
Proposed Method	0.81	<b>0.72</b>	<b>0.76</b>	0.87	<b>0.78</b>	<b>0.82</b>	0.84	<b>0.75</b>	<b>0.79</b>

### 4.3. Experiments on Natural Scene Datasets

As mentioned earlier in Section 4.1, the key steps are independent of the content of images. In other words, the steps proposed do not use any specific features of license plates, such as license plates that are in a horizontal direction and part of a vehicle, rather the proposed method considers license plates as one of the text types such as those in natural scene images. Therefore, in order to test the effectiveness of the proposed method on natural scene images, we conducted experiments on standard datasets, namely, ICDAR 2015, MSRA-TD-500, ICDAR 2017-MLT, MSCOCO, Total-Text and CTW1500. Qualitative results of the proposed method for the above datasets are shown in Fig.14, where one can see the proposed method detects well in images from different/diversified datasets. It is noted from the sample images shown in Fig.14 that images of each dataset exhibit different complexities. This shows that the proposed method is capable of handling different complex images of natural scene datasets. Quantitative results of the proposed and existing methods for the datasets, namely, ICDAR 2015, MSRA-TD-500, ICDAR2017-MLT, MSCOCO, Total and CTW1500 are reported in Table 5-Table 10, respectively.



Fig.14. Qualitative results of the proposed method for different natural scene benchmark datasets.

It is observed from Table 5-Table 10 that the proposed method achieves the best recall and F-score as in the license plate datasets for ICDAR2015, MSRA-TD-500, ICDAR2017-MLT compared to the existing deep learning-based methods. This shows that the proposed method is capable of addressing challenges of those datasets. For MSCOCO, the proposed method scores the best at recall compared to the existing methods. In this case, the proposed method scores low precision compared to the existing method due to the presence of more false positives from background information. However, the proposed method does not score high for the Total and CTW1500 dataset. This is because these two datasets are very complex compared to other datasets in terms of background, text orientation and text size variation. Although the score is low compared to the existing methods, one can argue that the proposed method is competitive as it outperforms a few existing methods. The low score for the Total-Text and CTW1500 datasets by the proposed method cannot be considered the major weakness because the main aim of the proposed work is to address the challenges of night and day license plate images but not natural scene text images. The existing methods report the best results at recall for all the datasets compared to the proposed method. This is due to the scope and objective of the existing methods. In other words, since the target of the existing

methods is to achieve the best results for natural scene texts, as expected, the methods perform well. Overall, when we compare the results of night/day license plate datasets and several natural scene text datasets, the proposed method scores consistent results irrespective of the datasets. Thus, the proposed method is effective and useful.

When we consider time complexity of the proposed method, the steps such as window expansion in the augmentation for dominant pixels detection, and searching for GVF loop patterns for candidate components detection, require more computations. Due to this, the proposed method consumes slightly more average processing time compared to the existing methods. For instance, 12-15 seconds for processing each image. However, it is noted that processing time depends on several factors, such as system configuration, coding, platform, language and implementation. Furthermore, due to the availability of GPUs and higher-end systems for experimentation, the processing time is not an important factor unless we target real-time implementation. Therefore, this work does not focus on achieving efficiency, rather focusing on finding a solution to the complex problem.

Table 5. The performance of the proposed method on the ICDAR 2015 natural scene dataset

Methods	Precision	Recall	F-Score
Yao et. al [24]	0.72	0.58	0.64
Tian et. al. [26]	0.74	0.52	0.61
Zhang et. al. [25]	0.71	0.43	0.54
StradVision2 [46]	0.77	0.36	0.49
StradVision1 [46]	0.53	0.46	0.49
NJU [46]	0.70	0.36	0.47
AJOU [27]	0.47	0.46	0.47
Deep2Text-MO [28]	0.49	0.32	0.38
CNN-MSER [46]	0.34	0.34	0.34
EAST [9]	<b>0.83</b>	0.78	<b>0.80</b>
Proposed Method	0.67	<b>0.79</b>	0.73

Table 6. The performance of the proposed method on the MSRA-TD500 natural scene dataset

Methods	Precision	Recall	F-Score
Kang et. al. [29]	0.71	0.62	0.66
Zhang et. al. [25]	0.83	0.67	0.74
He et. al. [14]	0.77	0.7	0.74
EAST [9]	<b>0.87</b>	0.674	0.76
SegLink [10]	0.86	0.7	0.77
Wu et. al. [14]	0.77	0.78	0.77
PixelLink [7]	0.83	0.732	0.778
TextSnake [30]	0.832	0.739	0.783
RRD [14]	0.87	0.73	<b>0.79</b>
Proposed Method	0.76	<b>0.82</b>	<b>0.79</b>

Table 7. The performance of the proposed method on the ICDAR2017-MLT natural scene dataset

Methods	Precision	Recall	F-Score
linkage-ER-Flow [49]	0.44	0.25	0.32
TH-DL [49]	0.67	0.34	0.45
SARI FDU RRPV v1 [49]	0.71	0.55	0.62
SCUT_DLVClab1 [49]	<b>0.80</b>	0.54	0.64

Sensetime OCR [49]	0.56	0.69	0.62
IDST_CV [49]	0.31	0.26	0.28
Proposed Method	0.66	<b>0.79</b>	<b>0.72</b>

Table 8. The performance of the proposed method on the MSCOCO natural scene dataset

Methods	Precision	Recall	F-Score
SCUT_DLVClab [48]	0.31	0.62	0.42
SARI_FDU_RRPN [48]	0.33	0.63	0.43
UM [48]	0.47	0.65	0.55
Text_Detection_DL [48]	<b>0.60</b>	0.61	<b>0.61</b>
Proposed Method	0.52	<b>0.71</b>	0.60

Table 9. The performance of the proposed method on the Total-Text natural scene dataset

Methods	Precision	Recall	F-Score
EAST [9]	0.5	0.36	0.42
SegLink [10]	0.30	0.23	0.26
TextSnake [30]	<b>0.82</b>	<b>0.74</b>	<b>0.78</b>
Proposed Method	0.79	0.72	0.75

Table 10. The performance of the proposed method on the CTW1500 natural scene dataset

Methods	Precision	Recall	F-Score
EAST [9]	<b>0.78</b>	0.49	0.60
SegLink [10]	0.42	0.40	0.40
DMPNet [31]	0.69	0.56	0.62
CTD [32]	0.74	0.65	0.69
TextSnake [30]	0.67	<b>0.85</b>	<b>0.75</b>
Proposed Method	0.45	0.65	0.53

#### 4.4. Discussion

The proposed work consists of four key steps, namely, augmentation for dominant pixel detection, checking GVF symmetry for candidate pixel detection, checking GVF loop patterns for candidate components detection, and deep learning for false positive elimination. Of the four steps, the proposed method defines heuristics for terminating the expansion process in the case of dominant pixels detection, GVF symmetry checking for candidate pixels detection, and GVF loop pattern checking for candidate component detection. It is true that the heuristics have their own inherent limitation such as sensitivity to images affected by severe distortion, being too dark, and other illumination effects. As a result, for these situations, the aforementioned conditions may not work well as shown in the sample images in Fig. 15, where for the input-distorted day and overly-dark night images in (a), the conditions miss pixels which represent license plate numbers as shown in (b) for the day image and in (c) for the night image. Thus there is scope for replacing the conditions by classifiers or iterative procedures in the future.

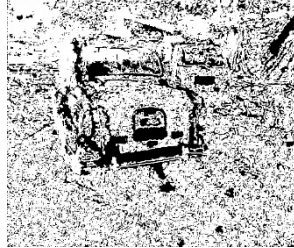




(a) Example of night and day images



Dominant pixels



Candidate pixels

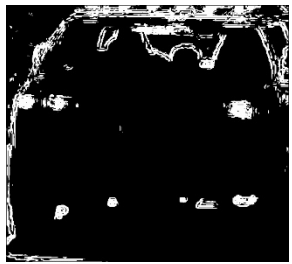


Candidate components



License plate detection

(b) The effect of the heuristics for distorted night image.



Dominant pixels



Candidate pixels



Candidate components



License plate detection

(c) The effect of the heuristics for a day image.

Fig. 15. Limitations of the heuristics used in the proposed work for distorted night and day images.

It is observed from the experimental results on the license plate datasets that for the night and day license plate datasets, the proposed method reports consistent results in terms of Recall and F-measure compared to the existing methods, while the existing methods report inconsistent results. The same conclusions can be drawn from the results on natural scene datasets. For the datasets of ICDAR2015, MSRA-TD-500, and ICDAR2017-MLT, the proposed method achieves the best Recall and F-measure compared to the existing methods, while for the MSCOCO and CTW1500 datasets, the performance of the proposed method degrades. This shows that the proposed method does not miss text instances in images, but due to background complexity, sometimes, the augmentation process classifies non-text as text for day images, and hence the precision is low compared to the existing methods. In the case of MSCOCO and CTW1500 datasets, it is expected that the stroke width varies arbitrarily, along with variations in the complexity of the backgrounds because of irregular-sized characters and texts in the images. Since the proposed method heavily relies on stroke width information, the proposed approach does not perform well. At the same time,

for images affected by severe blur due to car movements and very low contrast due to darkness, the proposed method fails to extract stroke width because stroke width is sensitive to blur and very low contrast.

It is also noted from the experimental results on the day and night license plate datasets that the methods developed for day images do not work well for night images and vice versa for the methods developed for night images. In this situation, if the methods of day images are trained with samples of night images, one can expect an improvement for night images. At the same time, if the methods of night images are trained with samples of day images, the performance of these methods improves for day images. But definitely, the improvement may not be significant because one cannot predict the complexity of the input images and the number of samples to be trained to achieve the best results for all types of images. It is evident from the results on the Night + Day datasets reported in Table 4, where the results are neither lower than night images nor higher than the day images for all the methods including those of night and day images. In addition, generating a large number of relevant, labeled data and ground truth for different applications is not an easy task. In order to reduce the dependency on the number of samples, and to make the method generic, the proposed work establishes the combination of hand-crafted features and deep learning. In this work, since the proposed method is trained with samples of night and day images, the proposed method reports consistent results for both night and day license plate images.

In the case of night license plate datasets, sometimes, one can expect full darkness, severe headlight effects, lights of other vehicles effecting the image and heavy blur due to vehicle movements as per the samples shown in Fig.16, whereby the proposed method underperforms. Therefore, there is scope for the proposed method to improve further by addressing the challenges of multiple vehicles in single images and other effects related to lights during the night. For images affected by severe illumination caused by headlights of many vehicles with street lights in a single image, if we develop an illumination model that can balance the lighting effects, one can expect improvements in the license plate detection results. Inspired by the methods [50, 51], where color and white balance are addressed to improve the segmentation of objects, in the same way, we can explore illumination models to balance light effects. However, this is beyond the scope of the proposed work and thus will be considered in the future.



Fig.16. Failure cases of the proposed method



## 5. Conclusions and Future Work

In this work, we have proposed a new method for addressing challenges of license plate detection in night and day license plate images. The proposed method employs the combination of color spaces and gradient information to tackle the causes of night and day images. The augmentation concept is introduced for detecting dominant pixels irrespective of night and day license plate images. Gradient Vector Flow is explored to find new patterns for text components in images for detecting candidate pixels as well as candidate components. Furthermore, recognition through a deep learning model is used for false positive elimination, merging text components and text detection. Experimental results on our night license plate dataset, the standard day license plate datasets and the benchmark datasets of natural scenes show that the proposed method is able to tackle the challenges of diversified datasets. It is evident from the quantitative results of the proposed and existing methods on different datasets that the proposed method outperforms the existing methods in terms of recall and F-score, and scores consistent results across the datasets. This is a distinct advantage of the proposed work. As noted in the Experimental Section, the proposed method reports poor results for precision and sometimes fails in the case of images affected by severe blur, full night conditions and headlight effects from multiple vehicles. To address these challenges, we shall propose a model for balancing illumination effects caused by headlights and full night conditions in the future. In addition, the methods that have been developed will be implemented in a real-time environment.

## Acknowledgements

This work was partially supported by a Faculty Grant: GPF014D-2019, University of Malaya, Malaysia and this work is supported by the Natural Science Foundation of China under Grant 61672273 and Grant 61832008, the Science Foundation for Distinguished Young Scholars of Jiangsu under Grant BK20160021.

## References

- [1] R. Panahi and I. Gholampour, "Accurate detection and recognition of dirty vehicle plate numbers for high speed applications", *IEEE Trans. ITS*, 18, pp 767-779, 2017.
- [2] C. H. Lin, Y. S. Lin and W. C. Liu, "An efficient license plate recognition system using convolutional neural networks", *In Proc. ICASI*, pp 224-227, 2018.
- [3] S. Du, M. Ibrahim, M. Shehata and W. Badway, "Automatic license plate recognition (ALPR): A state-of-the-art Review", *IEEE Trans. CSVT*, 23, pp 311-325, 2013.
- [4] H. Li, P. Wang and C. Shen, "Toward End to end car license plate detection and recognition with deep neural networks", *IEEE Trans. ITS*, 20, pp 1126-1136, 2019.
- [5] K. S. Raghunandan, P. Shivakumara, H. A. Jalab, R. W. Ibrahim, G. H. Kumar, U. Pal and T. Lu, "Riesz fractional based model for enhancing license plate detection and recognition", *IEEE Trans. CSVT*, 28, pp 2276-2288, 2018.
- [6] M. S. A. Shemarry, Y. Li and S. Abdulla, "Ensemble of adaboost cascades of 3L-LBPs classifiers for license plated detection with low quality images, *ESWA*, 92, pp 216-235, 2018.
- [7] D. Deng, H. Liu, X. Li and D. Cai, "PixelLink: Detecting scene text via instance segmentation", *In Proc. AAAI*, 2018.

- [8] A. Zamberletti, I. Gallo and L. Noce, "Augmented text character proposals and convolutional neural networks for text spotting from scene images", In Proc. ACPR, pp 196-200, 2015.
- [9] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, "East: an efficient and accurate scene text detector," in Proc. CVPR, pp. 2642–2651, 2017.
- [10] B. Shi, X. Bai, and S. Belongie, "Detecting oriented text in natural images by linking segments". In Proc. CVPR, pp. 3482–3490, 2017.
- [11] J. Ma, W. Shao, H. Ye, L. Wang, H. Wang, Y. Zheng, X. Xue, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals", IEEE Trans. MM, 20, pp 3111-3122, 2018.
- [12] X. Zhang, X. Gao and C. Tian, "Text detection in natural scene images based on color prior guided MSER", Neurocomputing, 307, pp 61-71, 2018.
- [13] W. He, X.-Y. Zhang, F. Yin, and C.-L. Liu, "Multi-Oriented and multi-lingual scene text detection with direct regression", IEEE Trans. IP, 27, pp 5406-5419, 2018.
- [14] C. Xue, S. Lu and W. Zhang, "MSR: multi-scale regression for scene text detection", In Proc. IJCAI, pp 989-995, 2019.
- [15] L. Deng, Y. Gong, Y. Lin, J. Shuai, X. Tu, Y. Zhang, Z. Ma and M. Xie, "Detecting Multi-Oriented Text with Corner-based Region Proposals", Neurocomputing, 21, pp 134-42, 2019.
- [16] Y. Xu, Y. Wang, W. Zhou, Y. Wang, Z. Yang and X. Bai, "TextField: Learning a deep direction field for irregular scene text detection", IEEE Trans. IP, 28, pp 5566-5579, 2019.
- [17] X. Zhao, C. Zhao, H. Guo, Y. Zhu, M. Tang and J. Wang, "Elite Loss For Scene Text Detection", Neurocomputing, 333, pp 284-291, 2019.
- [18] Y. Liu, L. Jin, S. Zhang, C. Luo and S. Zhang, "Curved scene text detection via transverse and longitudinal sequence connection", Pattern Recognition, pp 337-345, 2019.
- [19] D. Bazazian, R. Gomez, A. Nicolaou, L. Gomez, D. Karatazas and A. D. Bagdanov, "FAST: Facilitated and accurate scene text proposals through FCN guided pruning", Pattern Recognition Letters, 119, pp 112-120, 2019.
- [20] V. Khare, P. Shivakumara, R. Paramesran and M. Blumenstein, "Arbitrarily-oriented multi-lingual text detection in video", Multimedia Tools and Applications, Vol. 76, pp 16625-16655, 2017.
- [21] Z. Liu, W. Zhou and H. Li, "Scene text detection with fully convolutional neural networks", Multimedia Tools and Applications, Vol. 78, pp 18205-18227, 2019.
- [22] N. Gupta and A. S. Jalal, "A robust model for salient text detection in natural scene images using MSER feature detector and Grabcut", Multimedia Tools and Applications, Vol. 78, pp 10821-10835, 2019.
- [23] X. C. Yin, W. Y. Pei, J. Zhang, and H. W. Hao, "Multi-Orientation Scene Text Detection with Adaptive Clustering," IEEE Trans. PAMI, 37, pp. 1930-1937, 2015.
- [24] C. Yao, X. Bai, N. Sang, X. Zhou, S. Zhou, and Z. Cao, "Scene text detection via holistic, multi-channel prediction", arXiv preprint arXiv:1606.09002, 2016.
- [25] Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, and X. Bai, "Multi-oriented text detection with fully convolutional networks", In Proc. of CVPR, pp 4159-4167, 2016.
- [26] Z. Tian, W. Huang, T. He, P. He, and Y. Qiao, "Detecting text in natural image with connectionist text proposal network", In Proc. ECCV, pp 56–72, 2016.
- [27] H. Koo and D. H. Kim, "Scene text detection via connected component clustering and nontext filtering. IEEE Trans. on Image Processing, pp 2296–2305, 2013.
- [28] X. C. Yin, X. Yin, K. Huang, and H. Hao, "Robust text detection in natural scene images. IEEE Trans. on PAMI, pp 970–983, 2014.
- [29] L. Kang, Y. Li, and D. Doermann. "Orientation robust text line detection in natural images", In Proc. CVPR, pp 4034-4041, 2014.

- [30] S. Long, J. Ruan, W. Zhang, X. He, W. Wu, and C. Yao, "Textsnake: A flexible representation for detecting text of arbitrary shapes", In Proc. ECCV, pp , 19-35, 2018.
- [31] Y. Liu and L. Jin, "Deep matching prior network: Toward tighter multi-oriented text detection", In Proc. CVPR, pp 3454-3461, 2017.
- [32] Y. Liu, L. Jin, S. Zhang, S. Zhang, Detecting Curved Text in the Wild: New Dataset and New Solution, arXiv: 1712.02170.
- [33] M. Xue, P. Shivakumara, C. Zhang, T. Lu and U. Pal, "Curved text detection in blurred/non-blurred video/scene images", Multimedia Tools and Applications, pp 1-25, 2019.
- [34] P. Shivakumara, S. Roy, H. A. Jalab, R. W. Ibrahim, U. Pal, T. Lu. V. Khare, A. W. B. A. Wahab, "Fractional means based method for multi-oriented keyword spotting in video/scene/license plate images", Expert Systems with Applications, 118, pp 1-19, 2019
- [35] M. R. Asif, Q. Chun, S. Hussain, M. S. Fareed and S. Khan, "Multinational vehicle license plate detection in complex backgrounds", Journal of Visual Communication and Image Representation, 46, pp 176-186, 2017.
- [36] L. Xie, W. Ahmad, L. Jin, Y. Liu and S. Zhang, "A new CNN based method for multi-directional car license plate detection", IEEE Trans. ITS, 19, pp 507-517, 2018.
- [37] Y. Yuan, W. Zou, Y. Zhao, X. Wang, X. Hu and N. Komodakis, "A robust and efficient approach to license plate detection", IEEE Trans., IP, 26, pp 1102-1114, 2017.
- [38] X. Zhai, F. Bensaai and R. Sotudeh, "OCR-based neural network for ALPR", In Proc. ICMST, 2012.
- [39] A. H. Ahmed, M. Afifi, M. Korashy, E. K. William, M. A. El-sattar and Z. Hafez, "OCR systems for poor quality images using chain-code representation", In Proc. AISI, pp 151-161, 2015.
- [40] C. Zhang, P. Shivakumara, M. Xue, L. Zhu, T. Lu and U. Pal, "New fusion based enhancement for text detection in night video footage", In Proc. PCM, pp 46-56, 2018.
- [41] P. N. Chowdhury, P. Shivakumara, R. Raghavendra, U.Pal and M. Blumenstein, "a new U-net based license plate enhancement model in night and day images", In Proc. ACPR, pp 749-763, 2019.
- [42] P. Shivakumara, A. Konwer, A. Bhowmick, V. Khare, U. Pal and T. Lu, "A new GVF arrow pattern for character segmentation from double line license plate images", In Proc. ACPR, pp 782-787, 2017.
- [43] N. Boonsim and S. Prakoonwit, "Car make and model recognition under limited lighting conditions at night", Pattern Analysis and Applications, pp 1195-1207, 2017.
- [44] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, L.C. Chen, "MobileNetV2: Inverted Residuals and Linear Bottleneck", In Proc. CVPR, pp 4510-4520, 2018.
- [45] S. Tian, P. Shivakumara, T. Q. Phan, T. Lu and C. L. Tan, "Character shape restoration system through medial axis points in video", Neurocomputing, 161, pp 183-198, 2015.
- [46] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov, M. Iwamura, J. Matas, L. Neumann, V.R. Chandrasekhar, S. Lu, and F. Shafait, "ICDAR 2015 competition on robust reading. In Proc. ICDAR, pp. 1156-1160, 2015.
- [47] C. K. Ch'ng and C. S. Chan, "Total-text: A comprehensive dataset for scene text detection and recognition," In Proc. ICDAR, pp. 935-942, 2017.
- [48] A. Veit, T. Matera, L. Neumann, J. Matas and S. Belongie, "Coco-text: Dataset and benchmark for text detection and recognition in natural images", arXiv preprint arXiv:1601.07140, 2016.
- [49] N. Nayef, F. Yin, I. Bizid, H. Choi, Y. Feng, D. Karatzas, Z. Luo, U. Pal, C. Rigaud, J. Chazalon, et al. "ICDAR2017 robust reading challenge on multi-lingual scene text detection and script identification-RRC-MLT", In Proc. ICDAR, pp 1454-1459, 2017.
- [50] M. Afifi, B. Price, S. Cohen and M. S. Brown, "When color constancy goes wrong: Correcting improperly white balanced images", In Proc. CVPR, pp 1535-1544, 2019.

[51] M. Afifi and M. Brown, "What else can fool deep learning? Addressing color constancy errors on deep neural network performance", In Proc. ICCV, pp 243-252, 2019.