Text-Based Traffic Panels Detection using the Tiny YOLOv3 Algorithm

Saba Kheirinejad 1,*, Noushin Riahi 2, Reza Azmi 3
1Center for Ubiquitous Computing, University of Oulu, Oulu, Finland
2Department of Computer Engineering, Faculty of Engineering, Alzahra University, Tehran, Iran
3Department of Computer Engineering, Faculty of Engineering, Alzahra University, Tehran, Iran
*Corresponding author: Saba Kheirinejad, University of Oulu, Email: saba.kheirinejad@oulu.fi
Corresponding author ORCID: 0000-0002-3998-804X

ABSTRACT: Lately, traffic panel detection has been engrossed by academia and industry. This study proposes a new categorization method for traffic panels. The traffic panels are classified into three classes: symbol-based, text-based, and supplementary/additional traffic panels. Although few types of research have investigated text-based traffic panels, this type is considered in detail in this study. However, there are many challenges in this type of traffic panel, such as having different languages in different countries, their similarity with other text panels, and the lack of suitable quality datasets. The panels need to be detected first to obtain a reasonable accuracy in recognizing the text. Since there are few public text-based traffic panels datasets, this study gathered a novel dataset for the Persian text-based traffic panels all over the streets of Tehran-Iran. This dataset includes two collections of images. The first collection has 9294 images, and the latter has 3305 images. The latter dataset is more monotonous than the first one. Thus, the latter is utilized as the main dataset, and the first is used as an additional dataset. To this end, the algorithm uses the additional dataset for pre-training and the main datasets for training the network. The tiny YOLOv3 algorithm that is fast and has low complexity compared to the YOLOv3 is used for pre-training, training, and testing the data to examine the utility and advantages of the data. The K-fold cross-validation procedure is used to estimate the model’s skill on the new data. It achieves 0.973 for Precision, 0.945 for Recall, and 0.955 for F measure.


1. Introduction

Countries define rules to guarantee a secure traffic system. All traffic users such as drivers and pedestrians need to obey these rules. Traffic panels/signs play a crucial role in the system. Drivers observe the panels and act based on the information that the panels provide. The drivers intentionally or unintentionally ignore the traffic panels in different circumstances (once a vehicle has high speed or the driver is distracted by something), which can cause horrible accidents and misfortune. The definite dependence on the mentally and physically vulnerable human is hazardous. To this end, vehicles armed with progressive technologies play a significant role in the security of the traffic system and protecting many human lives. Detection of the traffic panels is considered a formidable technology in the intelligent transportation system. It has various uses including driver assistant systems, unmanned autonomous vehicles, and road panel keeping [1]. The traffic panels have been divided into two categories in the previous works: I) symbol-based and II) text-based traffic panel. This study classified the traffic panels into three classes: I) symbol-based, II) supplementary/additional, and III) text-based traffic panels.

1.1. Symbol-Based Traffic Panels

The first group utilizes specific and straightforward symbols to represent the concept of the panels. Symbol-based traffic panels are divided into several categories based on their form including rectangular, triangular, square, circular, pentagonal, and octagonal shapes, and their colors such as white, red, yellow, blue, and green. The shapes and colors with low variety determine the specific purpose of the panels, such as stop, pedestrian crossing, bicycle crossing, railway crossing, watching for children, speed limit, etc. The typical concepts of this type of traffic panel are danger, prohibition, commitment, warning, priority, end of prohibition, etc. Prohibition panels are in a circle form with a red margin, warning panels are triangular, and information panels are blue [1].
this type of panel [2]-[23]. Instances of symbol-based traffic panels can be seen in Figure 1.

![Instances of symbol-based traffic panels](image1.png)

**Figure 1**: Instances of symbol-based traffic panels.

1.2. Supplementary/Additional Traffic Panels

The traffic panels in this group complete the concept of symbol-based traffic panels. They contain ‘text,’ ‘arrow,’ ‘pictograph,’ and ‘a combination of the text and pictograph’ [24, 25, 26, 27, 28, 29, 30]. The shape of these panels is generally rectangular, their color is white, and they are found under the symbol-based traffic panels. There is a considerable variation of this type in Germany. However, there are a few numbers of this type in Iran. Since the number of traffic panels in this group is less than the number of traffic panels in two other groups, this group has attracted less attention. Instances of supplementary/additional traffic panels in Germany can be observed in Figure 2 and one instance in Iran can be observed respectively in Figure 3.

![Instances of supplementary traffic panels in Germany](image2.png)

**Figure 2**: Instances of supplementary traffic panels in Germany [24].

![Instances of supplementary/additional traffic panels in Iran](image3.png)

**Figure 3**: Instances of supplementary/additional traffic panels in Iran.  

1.3. Text-Based Traffic Panels

Text-based traffic panels include text and arrows. The texts deliver rich and valuable semantic information about the traffic system and have a fundamental function in the traffic system and intelligent transportation system. They illustrate the correct routes, warnings about possible dangers, and permission or prohibition of access to the road, etc. The automatic detection of these panels can alert the driver about the traffic and environment. Some systems can read the text using a synthesized voice or combine it with navigation systems like GPS. They play the assistant role and aim to increase the drivers’ attention to have a safe traffic environment. Fewer studies have been done in this category than symbol-based category [31, 32, 33, 34, 35, 36, 37, 38] [39, 40, 41, 42, 43, 44, 44]. The reason is a multitude of challenges which are presented in the following [1].

- There are not enough and suitable public datasets.
- The texts are written in different languages in different countries.
- Sometimes, the panels are not evident in various weather situations such as snowy, rainy, sunny, foggy, cloudy, etc.
- There are different text-based boards in the street that look like the text-based traffic panels such as billboards, store name boards, and advertisements on the vehicles.
- The color of the panels and texts fade due to the sunshine for a long time and reaction between the air and color.
- Some panels have been physically damaged.
- Some panels have been obstructed by the trees, vehicles, and other objects.

As far as we know, few benchmark datasets for text-based traffic panel detection are publicly available [45, 40]. Hence, this study gathered a novel public dataset all over the city of Tehran-Iran and called it ‘Persian Text-Based Traffic Panel Dataset’ [46]. As well as the general challenges mentioned above, there are other exclusive challenges for the Persian dataset: I) As shown in Figure 4, some traffic panels are not uniform from the perspective of color. In other words, they are a
combination of multiple colors. II) There are a few greens and white rectangular panels that only includes the symbols, and there is no text inside them, as shown in Figure 5. Since these panels look like text-based traffic panels, the network might make a mistake in detecting them. III) Persian text-based traffic panels include two languages, Persian and English. Therefore, the size of the texts is small, and the characters are unreadable from a relatively long distance.

1.4. Contribution

The contributions of this study are outlined as follows. This paper provides short and valuable literature on different types of traffic panels. Different traffic panels are categorized into a novel and comprehensive manner. The newest related works for detecting traffic panels in traditional and non-traditional methods are presented for each category. This study introduces a new dataset and elaborates on data collection and traffic panel detection techniques and challenges in detail. This work explains the way of evaluation using a state-of-the-art algorithm, the Tiny YOLOv3, to analyze the efficiency and benefits of the data. In the evaluation and results section, the algorithm's performance for the dataset has been extensively discussed. The results show that the algorithm provides 0.973 for Precision, 0.945 for Recall, and 0.955 for F-measure.

2. Related Work

This section discusses related works in different categories of traffic panels.

2.1. Symbol-Based Traffic Panel Detection

Detection methods for symbol-based traffic panels are generally divided into three sorts: I) colour-based methods, II) shape-based methods, and II) deep learning methods.

2.1.1. Color-Based Methods

The color-based methods usually use normalized color space such as RGB (Red, Green, Blue), YUV, and HSI (hue, saturation, and intensity), with specific thresholding. Segmentation means separating the traffic panels from the road background. Due to the damage to the color and shape of the traffic panels, it is not easy to distinguish its position from other outdoor images. Most segmentation techniques use color. Since the traffic panels are divided based on the color information from the background, it depends on the threshold of the input image in various color spaces. First, the segmentation is done using a different color range. Shape detection is done in the next step. Many researchers in this field face issues for white color and uneven lighting level [1]. In [47] red and green components' average is calculated for each pixel, and the G-R histogram is developed by displaying G-R's value. Many researchers use HSI color space because it understands the human understanding of RGB space well. While some researchers have found better results in the color space of YCbCr and YUV [48].

2.1.2. Shape-Based Methods

The particular shapes of traffic panels such as circles, triangles, and squares make them significant for distinguishing from other panels. For the shape-based methods, Hough transformation and different types of Hough transformation are used. Like detection of the general objects, a sliding window using categorization of the region is used to investigate the presence of traffic panels in the current window. Shape detection is an essential component of information in the recognition process of the traffic panels. Techniques such as connected components, dividing and merging edge detection, and clustering are utilized to compound pixels using color information-based similarity measurements.
to model the features at the detection stage. Attributes like aspect ratio, width, height, perimeter, and area of border-box, are extracted experimentally to obtain candidate bubbles. The resulting bubble is confirmed at this stage. There are several features that are extractable from the bubble that can be used in the classification of the shapes better [49]. Therefore, for each segmented bubble, right, left, bottom, and top DiB vectors are produced. Li et al. [50] utilized only right and left DiB vectors to detect the shape. The task in [51] is given by the union of distance vectors from the bubble’s center to the bubble’s outer edge, i.e., distance vector from the center of union. These feature vectors are utilized in the categorization of the shapes using a linear support vector machine. In [52] an alternative method uses the signature form of the connected components, which is computed directly through the bubble. Referring to the angle, the signature is defined the distance from the center to the object border. The absolute value of the fast Fourier transform is used for each shape signature to counteract the direction of the panel. The absolute value of discrete Fourier transform can be a good choice as well, but is not preferred due to computational complexity [53]. In [54], the Histogram of oriented gradients algorithm is developed using integrating the color information into the feature vector to improve performance. Since the road panels have high color contrast, some studies consider the pixels as features for shape definition [3].

2.1.3. Deep Learning Methods

Late advancements in deep learning techniques have inspired researchers to exploit neural networks to detect and recognize traffic panels. Contrary to the comparative achievement of detectors with hand-crafted features, most non-deep learning-based systems are not suitable to correctly detect a vast number of traffic panels [15]. Wu et al. converted the basic image into the grayscale image using support vector machines (SVM). They exploited convolutional neural networks (CNN) with fixed and learnable layers to detect the traffic panels [55]. Sun et al. utilized CNN and twin SVM hybrid model in a new study [18]. In [56] a traffic panel detection and recognition algorithm were presented using CNN. Target of the system included traffic panels, Chinese characters, English letters, and digits. The system uses a multi-task CNN as its base trained to obtain valuable features for classifying and localization of different traffic panels and texts. Ahmed et al. in a newer study proposed a CNN consists of a CNN-based challenge classifier [20]. In [57] for extracting the Region of Interest (RoI), the AdaBoost classifier and local binary pattern feature detector are combined. Cascaded CNNs are used to decrease negative samples of RoI for traffic panel recognition. Liu et al. also used cascade saccade machine learning network with hierarchical classes for traffic panel detection [22]. In [58] Abdi et al. presented a novel real-time method based on cascade deep learning and AR for a fast and precise framework for traffic panel recognition. It superimposes augmented virtual objects onto a real scene under all kinds of driving conditions, such as adverse weather circumstances. Kamal et al. merged the segmentation architectures Seg-Net and U-Net for detecting the traffic panels from video sequences [11]. Different Studies used CNN and mask R-CNN to detect the symbol-based traffic panel [12, 14]. Various studies used different versions of You Only Look Once (YOLO) algorithm, YOLOv3, YOLOv4, and YOLOv5 to recognize the traffic panels [16, 17, 23]. Wang et al. presented a new traffic panel recognition method via incorporating a lightweight superclass detector with a refinement classifier [59]. Nadeem et al. presented a new dataset in Pakistan and used transfer learning to recognize the traffic panels [21].

2.2. Additional/Supplementary Traffic Panel Detection

Since the supplementary or additional panels often appear below the symbol-based traffic panels, in most detection methods, first symbol-based panels are identified, then the interested region located below the symbol-based panels are specified, at last, the additional traffic panels in the RoI are searched. To this end, Nienhuser et al. presented an algorithm to detect additional panels besides the other panels. This method exploits Hough transform for lines and geometric constraints to find additional panels. Then uses SVM classification to verify the candidate shape [24]. Hamdoun et al. used MLP neural network to detect the rectangular shape in a region below speed-limit panels [30]. Puthon et al. presented a method to determine the shapes with ‘arrow,’ ‘pictogram,’ ‘text or mixed.’ Afterward, they classified the shapes with pyramid-HOG features [25]. Wenzel et al. utilized a Maximally Stable Extremal Regions (MSER) based approach to detect the additional panels [26]. In their next study, they found corner areas via aggregated channel features. Then, quadrangle generation and filtering technique were used to filter the variation of large aspect ratio for supplementary panels [28]. They introduced a complete pipeline to recognize the text by optical character recognition (OCR) in their earlier work. They assumed a specific additional panel, classified its layout, determined the bounding boxes of the content using regression, then used a multi-class classification step or applied a text
sequence classifier if necessary [27]. Weber et al. proposed CNNs to classify the road condition and recognize additional panels that display the validation of ‘when wet’ additional panels [29].

2.3. Text-Based Traffic Panel Detection

Few studies have been done in text-based traffic panels detection compared to symbol-based traffic panel. This study categorized methods of text-based traffic panels detection into two categories, I) the methods that extract the features in a traditional way, and II) the methods that extract the features using deep learning algorithms.

2.3.1. The Methods That Extract the Features in a Traditional Way

Wu et al. [35] used the traffic panels properties such as geometric constraints and color distribution to distinguish the traffic panels from the other objects. At last, they used the vertical plane criterion for traffic panel detection. Reina et al. [60] considered the traffic panels in Spain. They exploited the information of HSI color space and computation of the shape classification techniques for thresholding. They detected the blue pixels by applying an appropriate threshold and used achromatic decomposition to detect the white pixels. They labeled the connected components and considered all the blob candidates in a selection process and removed some of them based on their aspect ratio and size. Xavier et al. [32] found edges of the input image using a canny edge detector. They used Hough transform to find the straight lines of the detected edges. Then, they determined the candidate regions (top, right, and left of the road) that contain the traffic panels. They utilized HSV thresholding to detect the text-based traffic panels in the candidate regions. Finally, they vertically aligned the detected panels using the random sample consensus (RANSAC) algorithm and homography techniques. In [36], the images have been taken from Google Street View Service that provides 360° panoramic views with high resolution from different locations on various streets and roads worldwide. The authors used the text localization algorithm and detected the panel characters using MSER. They used MSER to determine traffic panels as rectangles with high text density. As plenty of traffic panels that have been studied in this paper have a blue background, a blue segmentation method was applied to detect the extensive rectangular regions. Greenhalgh et al. [37] specified the RoI of the text-based traffic panels by defining the vanishing point and sides of the road. Text-based traffic panels candidates were detected exploiting thresholding techniques in RoI by MSER and HSV. The number of candidates was decreased via temporal and contextual information at the end. Gonzalez et al. [38] detected the text-based traffic panels by Bag of Visual Words (BoVW) and color segmentation techniques. They applied the BoVW method on the specific parts of the images identified by the blue and white masking. Khodayari et al. [61] processed the input images in HSV space. Then, they detected the traffic panels using the fuzzy logic method. In the fuzzy images, the properties of each pixel, such as brightness and color are presented using linguistic values. Yellow, green, light green, white, dark red, and dark blue are examples of the linguistic values assigned to each pixel. Korhond et al. [45] presented a dataset including videos of traffic panels recorded by a moving vehicle in municipal regions. They categorized the video sequences based on their features in a way that researchers could select the piece of data that helps their research purpose.

2.3.2. The Methods That Extract the Features Using Deep Learning Algorithms

Rong et al. [31] proposed a new Cascaded Localization Network. They detected the candidate traffic panels in each frame on a set of continuous image frames using a set of features learned by CNN. They also collected a novel dataset for traffic guide panel to train and appraise the framework. Zhu et al. [62] collected a text-based traffic panels dataset containing English and Chinese traffic panels. They exploited a fully convolutional network to segment the candidate traffic panels. Luo et al. [42] proposed a new system for recognizing both symbol-based and text-based traffic panels. They applied MSER on grey and normalized RGB channels to extract the RoIs from each video frame. They trained a multi-task CNN with images labeled from street views and synthetic traffic panels and used it to recognize the panels. Eun et al. [33] used a region proposal network to detect Korean character candidates. Then, they classified Korean characters by a classification network. Jain et al. [34] applied MSER and OCR on given text-based traffic panels to recognize the texts. Peng et al. [43] presented a deep learning-based cascade detection model with two stages to detect the traffic panel text in natural scenes. They found the panels’ RoIs using an improved Single Shot Multi-Box Detector (SSD) network. Zhang et al. [44] solved the false detection and undetected by proposing a cascaded R-CNN to reach the multi-scale features in pyramids. All layers of the cascaded network combine the output bounding box of the previous layer for joint training; other than the first layer, this technique
chips for traffic panels detection. They also presented a multi-scale attention technique to get the weighted multi-scale features using softmax and dot-product, highlighting the traffic panel features and improving the detection accuracy by fining the features. Zhang et al. [63] proposed a new detection method called MSA_YOLOv3 to detect the small traffic panels precisely. They achieved data augmentation exploiting image mix-up technology. They introduced a multi-scale spatial pyramid pooling block into the Darknet53 network to help the network to learn object features more generally. Guo et al. [39] proposed an algorithm to detect the mixed horizontal-and-vertical-text traffic panel with the Chinese language in the street. To differentiate the traffic panels from other similar things in color in the complicated background like street scenes, the different red, green, and blue features were effectively mixed. Since the Chinese text lines are usually vertical and horizontal on text-based traffic panels, the presented technique created the text lines by the structural information and position of the characters. Boujemaa et al. [40] presented a novel public multi-task dataset for detecting text-based traffic panels. Bagi et al. [41] presented an end-to-end trainable deep neural network that can recognize multi-oriented text instances in adverse meteorological conditions.

3. The Tiny YOLOv3 Algorithm

Since the computer’s hardware is continuously growing, CNN-based deep learning methods have advanced fast and acquired remarkable results in computer vision and machine vision scope [64]. Traffic panel detection is considered a subset of the more general scope of object detection. Object detection is a relatively old subject. However, it is still a complex issue and an active scope for many researchers despite the endeavors that have been accomplished. Various object detection algorithms have been suggested so far. Currently, the most famous algorithms are SSD [65], R-CNN [66], and YOLO. YOLO is an algorithm that has become widespread for object detection recently. Joseph Redmon et al. have proposed its three versions (1-3) [67, 68, 69]. The fourth and fifth versions have been presented by [70] a [71], respectively. This algorithm solves object detection as a regression problem. It classifies the objects and gives their location as an output using an end-to-end network in one step. The noteworthy thing about this algorithm is its speed. It is one of the fastest algorithms so far. YOLOv3 uses the K-means clustering algorithm to automatically choose the most suitable anchors for the dataset. This algorithm is very complicated and demands complex hardware. It has 256 layers, and 53 layers out of 256 are convolution layers, as it utilizes the Darknet-53 as the backbone network. The tiny YOLOv3 is a smaller and faster version of the YOLOv3 algorithm. It has an architecture with low complexity and a more straightforward implementation for datasets with small sizes. This algorithm is used for pre-training, training, and testing the dataset. The tiny YOLOv3 algorithm used in this study has 46 total layers, including 13 Convolution, 11 Leaky Relu, 11 Batch Normalization, 6 Max Pooling, 3 Input, 1 Concatenate, and 1 Upsampling layer. In Table 1, the features of the tiny YOLOv3 architecture are summarized.

Table 1: Features of the tiny YOLOv3 architecture.

<table>
<thead>
<tr>
<th>Name of layers</th>
<th>Number of layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Convolution</td>
<td>13</td>
</tr>
<tr>
<td>Batch Normalization</td>
<td>11</td>
</tr>
<tr>
<td>Leaky Relu</td>
<td>11</td>
</tr>
<tr>
<td>Max Pooling</td>
<td>6</td>
</tr>
<tr>
<td>Input</td>
<td>3</td>
</tr>
<tr>
<td>Upsampling</td>
<td>1</td>
</tr>
<tr>
<td>Concatenate</td>
<td>1</td>
</tr>
</tbody>
</table>

4. Dataset Collection

Before getting into the details of collecting the dataset, the characteristics of the Persian text-based traffic panels in Iran are explained. These traffic panels express a particular meaning according to their shape and color. The concepts of the colors for this type of traffic panel are as follows:

- Green: Route guide for highways and religious places
- Blue: Route guide for freeways and service guide
- Black and white: Route guide for the other ways
- Yellow: General warnings and warnings for construction or repair operations
- Orange: Guide for administrative, training, and service areas
- Brown: Guide for recreational, cultural, and tourism areas

The concept of the shapes in the text-based traffic panels are as follows:

- Horizontal rectangle: Warning for repairing and maintenance
- Vertical rectangle: Guide for imperative and service panels
- Flag rectangle: Guide for routes

Since few public datasets focus on text-based traffic panels, this study collected a novel challenging dataset containing traffic panels with the Persian and English text called ‘Persian Text-Based Traffic Panels.’ In this dataset, the images have been taken using diverse smartphone cameras with various specifications by crowdsourcing.
method. Most images have been taken all over the city of Tehran in Iran, more on highways and streets. Two collections of images were collected. First, 4000 images were collected in different situations. Instances of these images can be witnessed in Figure 6. The images were augmented to 9294, followed by labeling. The tiny YOLOv3 algorithm was used for training and testing to examine the utility and advantages of the data. Nevertheless, this dataset did not attain suitable performance despite the positive expectation. The reason might be numerous text-based traffic panels in different situations compared to the total number of images (9294). In other words, there were many challenges in that dataset. It was compared to the dataset (an available part of the dataset) that has been used in [31]. Since most investigated text-based traffic panels are green and somewhat look like each other, this dataset is nearly monotonous and does not have many challenges. Hence, training and testing a model with this dataset is more straightforward than our dataset, gaining more accurate results. Thus, 1500 new images were taken as the main dataset. These images were more uniform than the images in the additional dataset. The images were augmented to 3305 and were labeled. This dataset was called the ‘main dataset’ and was utilized for training the algorithm. It is worth noting that the previous dataset was used to pretrain the algorithm and was called ‘additional dataset.’

4.1. Additional Dataset Collection

This dataset contains 4000 images with 12 Gigabyte sizes. Since different persons and cameras have taken the images, they are in both vertical and horizontal shapes, and their sizes are different. In other words, the images are not uniform. The images needed to be preprocessed to get uniform for use in the network. The desired size of the input images for the network is 416*416. If the images were directly resized to 416*416, the objects within the image would lose their original form. To this end, the cropping technique was used, and the images were converted to square form with the same length and width. Then, a bicubic interpolation method was applied to make the images monotonous.

The cropping strategy has a pair of advantages: I) data augmentation and II) making the images uniform. To augment the data, the images were cropped three times, from the left side, right side, and both sides. Using this technique, the total number of images increases, and no pixels are discarded from the original images, i.e., all the pixels were optimally used. After cropping each image, three images are produced from one original image. However, in some images, cropping causes the exit of some traffic panels from the cropped images. Thus, the images with no traffic panels need to be removed, which made the number of the final images not precisely three times the number of the original images. Therefore, the total number of augmented images is 9294. Finally, a
bicubic interpolation technique was applied to derive a uniform dataset with 416*416 size.

4.2. Main Dataset Collection

Due to the additional dataset challenges, a new and more uniform dataset was collected. 1500 images were collected and augmented to 3305 using the exact technique utilized for the additional dataset. Instances from the main dataset can be observed in Figure 7. In the new dataset, the panels showing the name of the alleys and streets were ignored. Because they are smaller than the usual traffic panels, and their texts are unclear. The green and white text-based traffic panels were used more. The images that a big part of them have been blocked with the trees, cars, and the other objects were not also used.

After data collection and applying the preprocessing steps, the images were labeled. Before explaining the labeling process, the predicted bounding box and ground-truth concepts need to be defined. The predicted bounding box is the bounding box predicted by the network. The ground-truth bounding box is the exact location of the defined traffic panel in the images. As shown in Figure 8 the predicted bounding box and ground-truth bounding box have been depicted in red and green. The predicted bounding boxes that overlap more with the ground-truth bounding boxes are more precious. Ground-truth bounding boxes are usually determined in two forms: I) the coordinates of the two corner points connected by diameter, and II) the coordinates of one corner and the length and width of the bounding box. This study uses the first one. The authors developed a code, after running the code, the dataset’s images are displayed one by one. As it can be seen in Figure 9, the user needs to click twice (one clicks for left up corner and another one for right down corner) for each panel, in such a way that the bounding box covers the whole panel.
between ground-truth and predicted bounding box over the area of the union between ground-truth and predicted bounding box. This study considers the IoU threshold 0.5. The Precision, Recall, and $F_{\text{measure}}$ determining the accuracy of the algorithm are described in the following.

True Positive: TP  
False Positive: FP  
False Negative: FN

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

$$F_{\text{measure}} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3)$$

This study uses K-Fold cross-validation technique with K=5 to evaluate the algorithm’s performance. The results are illustrated in Table 4. The evaluation outcomes of Fold1 for various thresholds of IoU are depicted in Table 5. As shown, once the IoU threshold boost, Precision, Recall, and $F_{\text{measure}}$ drop. The comparison of this method with other methods for text-based traffic panel detection can be observed in Table 6.

Table 2: Parameters’ value of the network.

<table>
<thead>
<tr>
<th>Input image size</th>
<th>416 * 416</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchor number</td>
<td>6</td>
</tr>
<tr>
<td>Class number</td>
<td>1, traffic panel class</td>
</tr>
<tr>
<td>Optimized function</td>
<td>Adam</td>
</tr>
<tr>
<td>Learning rate</td>
<td>0.001</td>
</tr>
<tr>
<td>Batch size</td>
<td>8</td>
</tr>
<tr>
<td>Epoch</td>
<td>100</td>
</tr>
<tr>
<td>Patience</td>
<td>5</td>
</tr>
<tr>
<td>Graph card</td>
<td>NVIDIA GeForce GTX1070</td>
</tr>
</tbody>
</table>

Table 3: Characteristics of the additional and main dataset.

<table>
<thead>
<tr>
<th>Number of images in</th>
<th>Additional dataset</th>
<th>Main dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total collection</td>
<td>4000</td>
<td>1500</td>
</tr>
<tr>
<td>Total collection after data augmentation</td>
<td>9294</td>
<td>3305</td>
</tr>
<tr>
<td>Train collection</td>
<td>7435</td>
<td>2644</td>
</tr>
<tr>
<td>Validation collection</td>
<td>743</td>
<td>264</td>
</tr>
<tr>
<td>Test collection</td>
<td>1859</td>
<td>661</td>
</tr>
</tbody>
</table>

Table 4: Precision, Recall, and $F_{\text{measure}}$ for each five folds in cross validation.

<table>
<thead>
<tr>
<th>Test subset</th>
<th>Image number</th>
<th>Traffic panels number</th>
<th>True Positive number</th>
<th>False Positive number</th>
<th>False Negative number</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_{\text{measure}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fold1</td>
<td>661</td>
<td>871</td>
<td>823</td>
<td>24</td>
<td>48</td>
<td>0.971</td>
<td>0.944</td>
<td>0.957</td>
</tr>
<tr>
<td>Fold2</td>
<td>661</td>
<td>880</td>
<td>829</td>
<td>27</td>
<td>51</td>
<td>0.968</td>
<td>0.942</td>
<td>0.954</td>
</tr>
<tr>
<td>Fold3</td>
<td>661</td>
<td>865</td>
<td>819</td>
<td>46</td>
<td>77</td>
<td>0.977</td>
<td>0.946</td>
<td>0.960</td>
</tr>
<tr>
<td>Fold4</td>
<td>661</td>
<td>850</td>
<td>807</td>
<td>17</td>
<td>43</td>
<td>0.979</td>
<td>0.949</td>
<td>0.963</td>
</tr>
<tr>
<td>Fold5</td>
<td>661</td>
<td>875</td>
<td>827</td>
<td>25</td>
<td>48</td>
<td>0.970</td>
<td>0.945</td>
<td>0.944</td>
</tr>
<tr>
<td>Average</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.973</td>
<td>0.945</td>
<td>0.955</td>
</tr>
</tbody>
</table>

Table 5: Precision, Recall, and $F_{\text{measure}}$ for various IoU in Fold1.

<table>
<thead>
<tr>
<th>IoU</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_{\text{measure}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0.5</td>
<td>0.971</td>
<td>0.944</td>
<td>0.957</td>
</tr>
<tr>
<td>&gt;0.6</td>
<td>0.955</td>
<td>0.940</td>
<td>0.950</td>
</tr>
<tr>
<td>&gt;0.7</td>
<td>0.818</td>
<td>0.935</td>
<td>0.873</td>
</tr>
<tr>
<td>&gt;0.8</td>
<td>0.509</td>
<td>0.900</td>
<td>0.650</td>
</tr>
</tbody>
</table>

Table 6: Comparison of different methods for text-based traffic panel detection.

<table>
<thead>
<tr>
<th>Study</th>
<th>Precision</th>
<th>Recall</th>
<th>$F_{\text{measure}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rong et al. [31]</td>
<td>0.73</td>
<td>0.64</td>
<td>0.68</td>
</tr>
<tr>
<td>Greenhalgh et al. [37]</td>
<td>0.96</td>
<td>0.90</td>
<td>0.93</td>
</tr>
<tr>
<td>Zhu et al. [62]</td>
<td>0.93</td>
<td>0.94</td>
<td>0.93</td>
</tr>
<tr>
<td>Peng et al. [43]</td>
<td>0.93</td>
<td>0.91</td>
<td>0.92</td>
</tr>
<tr>
<td>This study</td>
<td>0.97</td>
<td>0.94</td>
<td>0.95</td>
</tr>
</tbody>
</table>

6. Discussion

Good accuracy in the high IoU thresholds is crucial as the detected text-based traffic panels may not contain the full text. The used algorithm in this study has some flaws mentioned in the following. As shown in Figure 10, the predicted bounding boxes do not contain the full text though IoU is higher than 0.5. Since the next step after the text-based traffic panel detection is the recognition of the texts, the predicted bounding boxes need to contain the entire text to deliver the concept. As shown in Figure 10, predicted bounding boxes that do not include the full text.

Figure 11, despite the clarity of some panels in specific sizes, the network cannot detect them. The reason is anchors’ size chosen by the algorithm at the beginning. As depicted in Figure 12, the panels that were wrongly detected as the traffic panels look like some text-based traffic panels in size and background. Although the Tiny YOLOv3 has a few flaws, it is a robust algorithm and can detect traffic panels in many challenging situations. It
does not wrongly detect the panels that contain several symbol-based panels (indeed, they are considered symbol-based traffic panels) and look like the text-based traffic panels, as shown in Figure 13. As depicted in Figure 14, other than a few cases mentioned as the algorithm’s flaws, it does not incorrectly detect the other similar objects to the text-based traffic panels such as the advertisements and billboards on the sides of the roads and vehicles, store name board, etc. It can accurately detect the text-based traffic panels in the complicated backgrounds as shown in Figure 15. It can correctly detect the lopsided text-based traffic panels, as shown in Figure 16. As shown in Figure 17, it can detect the text-based traffic panels partially blocked by objects, including vehicles, other panels, trees, etc.

7. Conclusion

This study proposed a new categorization method for traffic panels. It classified the traffic panels into three classes such as symbol-based, text-based, and supplementary/additional traffic panel. A novel text-based traffic panel dataset named ‘Persian text-based traffic panel dataset’ was introduced. It includes two collections of images. The first is a collection of 9294 images utilized as an additional dataset for pre-training the network. The latter is the main dataset, including 3305 images, and was utilized for training and testing the network.

The tiny YOLOv3 algorithm was exploited for pre-training, training, and testing the data to examine the utility and advantages of the data. The K-fold cross-validation procedure was used to estimate the model’s skill on the new data. It achieves 0.973 for Precision, 0.945 for Recall, and 0.955 for F measure. Since the next step of text-based traffic panel detection is recognizing the text on the traffic panel, the authors plan to enhance text-based traffic panel detection accuracy for IoU thresholds greater than 0.5 to focus on recognizing the texts in the dataset as the future study.
Figure 16: Instances of the correctly detected lopsided text-based traffic panels.

Figure 17: Instances of the partially blocked text-based traffic panels that have been correctly detected.

References


S. Kheirinejad et al., Text-Based Traffic Panel Detection


Copyright: This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

Saba Kheirinejad received the M.Sc. degree in Artificial Intelligence and Robotics from Alzahra University, Tehran, Iran. She is currently pursuing a Ph.D. degree with the Centre for Ubiquitous Computing, University of Oulu, Finland. Her current research interests include ubiquitous computing, human-computer interaction, and digital health.

Noushin Riahi received the B.S. degree from Isfahan University of Technology in 1986 and the M.S. and PhD. degree in Electrical and Electronic Engineering from Sharif University of Technology, Tehran, Iran in 1990 and 1998. Her research interests include speech and sound processing, opinion mining, recommender systems, social network processing, and biological signal processing. Currently, she is associate professor at Alzahra University, computer group. Topics taught by her include speech processing, advance computer architecture, digital signal processing, natural language processing, signals and systems theory, microprocessors, digital design, etc.

Dr. Reza Azmi received his BS degree in Electrical Engineering from Amirkabir University of Technology (Tehran polytechnic) in 1990 and his MS and PHD degrees in electrical and computer Engineering from Tarbiat Modares University, Tehran, Iran in 1993 and 1999 respectively. Since 2001 he has joined computer department of Faculty of engineering at Alzahra University. He was member of multimedia systems working group in ITRC, Optical Character Recognition working group in supreme council of information and communication technology, he was manager and technical member of many industrial projects and four US international patents. Dr. Azmi is founder and co-founder in three startup companies. He has received numerous awards for being selected as a top industrial researcher.