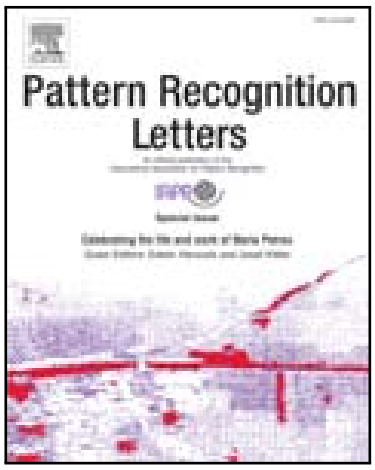


Graph Attention Network for Detecting License Plates in Crowded Street Scenes

Pinaki Nath Chowdhury , Palaiahnakote Shivakumara , Swati Kanchan , Ramachandra Raghavendra , Umapada Pal , Tong Lu , Daniel Lopresti

PII: S0167-8655(20)30354-8
DOI: https://doi.org/10.1016/j.patrec.2020.09.018
Reference: PATREC 8037

To appear in: *Pattern Recognition Letters*

Received date: 22 April 2020
Revised date: 14 August 2020
Accepted date: 20 September 2020

Please cite this article as: Pinaki Nath Chowdhury , Palaiahnakote Shivakumara , Swati Kanchan , Ramachandra Raghavendra , Umapada Pal , Tong Lu , Daniel Lopresti , Graph Attention Network for Detecting License Plates in Crowded Street Scenes, *Pattern Recognition Letters* (2020), doi: https://doi.org/10.1016/j.patrec.2020.09.018

Highlights

- A new problem of multiple vehicles license plate detection is addressed.
- Proposed a new framework that combines ResNet and GAT layers in a new way
- The proposed method outperforms the existing methods.

Pattern Recognition Letters
journal homepage: www.elsevier.com

Graph Attention Network for Detecting License Plates in Crowded Street Scenes

[1]Pinaki Nath Chowdhury, [2]Palaiahnakote Shivakumara, [1]Swati Kanchan, [3]Ramachandra Raghavendra, [1]Umapada Pal, [4]Tong Lu and [5]Daniel Lopresti

[1]*Computer Vision and Pattern Recognition Unit, Indian Statistical Institute, Kolkata, India. Email: pinakinathc@gmail.com, umapada@isical.ac.in, swatikanchan070@gmail.com.*
[2]*Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia. Email: shiva@um.edu.my*
[3]*Faculty of Information Technology and Electrical Engineering, IIK, NTNU, Norway, raghavendra.ramachandra@ntnu.no*
[4]*National Key Lab for Novel Software Technology, Nanjing University, Nanjing, China. Email: lutong@nju.edu.cn*
[5]*Computer Science & Engineering, Lehigh University, Bethlehem, PA, USA. Email: lopresti@cse.lehigh.edu*

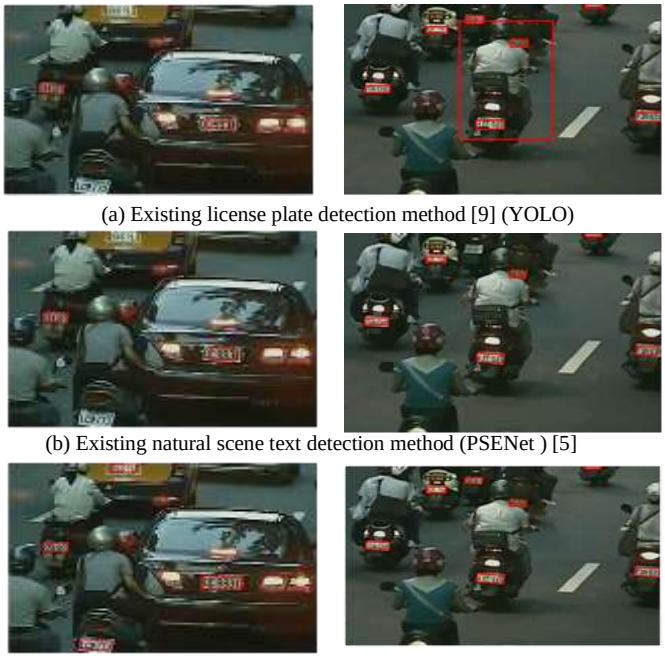
ABSTRACT

Detecting multiple license plate numbers in crowded street scenes is challenging and requires the attention of researchers. In contrast to existing methods that focus on images that are not crowded with vehicles, in this work we aim at situations that are common and complex, for example, in city environments where numerous vehicles of different types like cars, trucks, motorbike etc. may present in a single image. In such cases, one can expect large variations in license plates in terms of quality, backgrounds, and various forms of occlusion. To address these challenges, we explore Adaptive Progressive Scale Expansion based Graph Attention Network (APSEGAT). This approach extracts local information which represents the license plates irrespective of vehicle types and numbers because it works at the pixel level in a progressive way, and identifies the dominant information in the image. This may include other parts of vehicles, drivers and pedestrians, and various other background objects. To overcome this problem, we integrate concepts of graph attention networks with progressive scale expansion networks. For evaluating the proposed method, we use our own dataset, named as AMLPR, which contains images captured in different crowded street scenes in different time span, and the benchmark dataset namely, UFPR-ALPR, which provides images of a single vehicle, and another benchmark dataset called, UCSD, which contains images of cars with different orientations. Experimental results on these datasets show that the method outperforms existing methods and is effective in detecting license plate numbers in crowded street scenes.

1. Introduction

Traffic congestion on busy streets in cities creates a challenging environment for computer vision and pattern recognition techniques. Often there are many vehicles in a street without much space between two vehicles. In this situation, captured images will contain many different vehicles at different angles and positions [1]. The same is true of crowded parking lots, and also toll booth gates on highways. These situations provide the motivation for the work reported in this paper. There are several methods developed in the past for license plate number detection by addressing the issues of complex background, multi-lingual, orientation, dirty plates, and noisy plates [2,3,4]. Most methods consider images having a single vehicle. However, none of the methods focuses on images of crowded street scene. Therefore, existing approaches may not work well when there are many vehicles in an image. Occlusion, for example, can cause serious problems for detection. In addition, varying degrees of focus due to the different distances of license plate numbers in an image may have an impact on detection. Hence, it is useful to develop methods that can detect license plate number in crowded street scenes, irrespective of the number of vehicles and their types.

Similarly, methods have been developed for text detection in natural scene images that can handle multiple instances in a single image [5, 6]. However, the existing methods of natural scene images which are dependent on the text strings having minimum length and sharing common properties, may not work well for the application we have described.



(a) Existing license plate detection method [9] (YOLO)

(b) Existing natural scene text detection method (PSENet) [5]

(c) The proposed method

Fig.1. Examples of detecting license plate in the images of dense vehicles in a single image by the existing and the proposed methods.

As shown in Fig. 1, we can see that the YOLO architecture [7], which is used widely by past methods [8, 9,10] for license plate number detection, detects almost all the license plate

numbers but at the same time, it produces false positives as shown in Fig. 1(a) for the right side image. Likewise, the natural scene text detection method [5] which uses Progressive Scalable Expansion Network (PSENet) fails to detect all of the license plate numbers in Fig. 1(b). However, our method correctly detects all of the license plate numbers in these two example images, as shown in Fig. 1(c). Therefore, one can conclude that these two examples illustrate the kinds of limitation of the existing methods for detecting license plate number in the images of the crowded street scenes. Therefore, this work focuses on developing a unified method based on ResNet and GAT (Graph Attention Network) for detecting multiple license plate numbers in complex environment.

The contributions of our work are as follows:

(a) The way we integrate the strength of the ResNet and GAT as a unified model is the key contribution.

(b) One of the key advantages of the proposed method is the detection of multiple license plate numbers of different vehicles in a single image in a complex environment.

(c) In addition, since the problem is a complex and open issue especially for detecting license plates in the crowded scene, the proposed model can be a baseline method for further investigation.

2. Related Work

For the past decade, several powerful methods are developed for license plate number detection and recognition by addressing many challenges, such as complex background, plates affected by dust, orientation, noise, blur etc. Overall, the existing methods can be broadly categorized as two sub-categories.

2.1. The Methods for Single License Plate Detection

Hendry et al. [8] proposed an approach for license plate number detection based on the architecture called You Only Look Once (YOLO-7). The method uses a sliding window process for detection in the images. The method works well for different situations, like rainy background, darkness and dimness, and different hues and saturation of images. However, the approach is sensitive to low contrast and low-resolution images. Kessentiini et al. [9] proposed a method for multi-lingual license plate detection in the images. The approach uses YOLO architecture for extracting features in two stages: (i) license plate detection and (ii) for license plate recognition. However, the approach does not work well for the images affected by uneven illumination and blur.

Li et al. [11] proposed a method for license plate detection based on deep learning model. The method aims at unifying a model for both detection and recognition. However, the method is not invariant to the orientation of license plate images. Liu and Chang [12] proposed a license plate detection method with application to large surveillance applications. The approach uses cascaded color space transformation of pixel detector and the cascaded color Haar-like detector for license plate detection. However, the method is computationally expensive. Min et al. [10] proposed a method for vehicle license plate location based on YOLO architecture. The approach targets the images affected by different weather conditions and viewpoint changes. The method uses k-means clustering and YOLO architectures for detecting the license plate in the image. However, the success of license plate detection depends on the success of region detection. Moreno et al. [13] proposed a scale adaptive license plate detection method based on a part-based approach. The target of the method is to detect license plates of different sizes, which works based on boosting algorithm. However, the approach is sensitive to illumination effect.

Omar et al. [14] proposed a method for license plate detection based on cascaded deep learning. The approach employs several pre-processing steps to enhance the fine details of the license plate region. Then deep semantic segmentation network is used for detecting license plate regions in the images. Since the method uses several pre-processing steps, the approach may not be robust to images of different situations. Selmi et al. [2] proposed a method for license plate detection based on a deep neural network by targeting multi-lingual and multi-sized license plate images. The approach uses mask region convolutional neural networks for license plate detection in the images. The approach is considered as computationally expensive. Shemarry et al. [15] proposed a method to define texture descriptor for detecting license plate in the images of difficult conditions. The approach employs Gaussian filter and histogram equalization for enhancing the license plate regions. Then it uses a multi-level extended local binary pattern for license plate detection in the images. However, the approach generates several false positives for the images of different conditions.

Silva et al. [16] proposed a license plate detection method in real-time environment based on deep convolutional neural networks. The approach aims at identification of vehicle and then license plate detection to achieve better results. To achieve this, the method adapts convolutional neural networks. The main weakness of this method is that the accuracy depends on the success of vehicle identification step. Shivakumara et al. [6] proposed a method that considers license plate number detection as text detection in natural scene images. Therefore, the approach employs fractional-means method for license plate detection in the images. Since the main goal of the method is to detect in the different images but not the license plate detection, the results are not consistent for the images of different situations.

It is noted from the review of the above methods that the approaches address several challenges of license plate detection in the images of different conditions. However, none of the methods addresses the challenges of the images containing dense vehicles captured in crowded street scenes. Most of the approaches consider the images having a single vehicle in a single image.

2.2. The Methods for Multiple License Plate Detection

There are methods for detecting multiple license plates in a single image by targeting vehicles on multilane. For instance, Asif et al. [1] proposed a method for license plate detection in a multi-lane environment. The approach firstly detects tail lights based on color information and then edge information is used for license plate detection. As long as the method detects tail lights, the approach works well but sometimes, when the tail lights are damaged or obstructed, the methods may not work well. Bhargav and Deshpande [17] proposed a method for locating multiple license plates based on color clustering and filtering techniques. The method focuses on license plates of different countries but not multiple vehicles in a single image. Also, the proposed connected component-based method may not be robust to the images of different conditions.

Luo et al. [18] proposed a method for multiple Chinese vehicle license plate detection in the images based on Single Shot Multi-Box Detector (SSD). The approach uses corner points detection and character contour detection for license plate localization. The method may work well for high-quality images but not for low-quality images. Menon and Omman [19] proposed a method for multiple license plate detection in a single

image. The approach extracts geometrical features of the license plate, such as shapes, contours etc. for detection. However, the extracted features are good for high quality and contrast images but not for the images with complex background and poor quality. Zhou et al. [20] proposed a method for detecting multiple license plate locations in the images based on cascaded and convolutional neural network. However, the scope of the method is limited to particular types of images.

In summary, it is observed from the above review that there are methods for the images having many cars but not the images of crowded scenes. Therefore, none of the methods used crowded scene images for detecting the license plate number. When the image contains especially dense motorbikes or motorcycles along with the cars, trucks, drivers and pedestrian, partial occlusion and varying degree of focus for the vehicles in the images make the license plate number detection more complex. Therefore, there is a need for developing a method for detecting license plate number in crowded scenes.

Hence, in this work, to overcome the problems of small and poor-quality license plate number caused by motorcycle and to distinguish license plate number from other information in the image, we propose Adaptive Progressive Scalable Expansion based Graph Attention Network (APSEGAT). The motivation to adapt PSENet is that it constructs uniqueness by collecting information from pixel to component level progressively, which is essential to cope with tiny font, degradations and poor quality [5]. We believe that the progressively collected information provide unique information for the license plate number irrespective of challenges. Since the considered problem is complex and there are high chances of losing discriminative power of the feature extraction, inspired by Graph Attention Network (GAT), which helps us to find relevant information through inductive learning [21], we integrate GAT with PSENet for detecting license plate in the images containing dense vehicles captured in crowded street scenes, which results in APSEGAT.

The organization of the proposed work is as follows. Section 3.1 presents a discussion on the proposed combination of PSENet and GAT. The procedure for developing end-to-end learning of the proposed model is discussed in Section 3.2. Experimental results are presented in Section 4. Finally, conclusion and future work are briefed in Section 5.

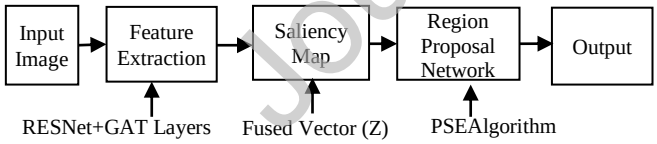


Fig. 2. Block diagram of the proposed method.

3. Proposed Methodology

It is noted that the variation among pixel values of the license plate numbers across vehicles is almost uniform. In other words, although the pixel values differ from one license plate number to another in the same image, the variation in the pixel values is almost constant compared to the pixels value of non-license plate regions. This is the common property of license plate number irrespective of vehicle. To extract such observation, we propose to integrate the PSENet and GAT models as a single model for solving such a complex problem of license plate number detection in crowded scenes.

For the input image, the proposed method uses ResNet to derive Progressive Scalable Expansion Network (PSENet) for extracting features because as mentioned in the previous section, PSENet helps us to accumulate local information from pixel to component level in a progressive way. However, PSENet alone is not enough to separate license plate numbers from non-license plate numbers in the images. Therefore, to overcome this problem, the proposed method integrates Fwith the PSENet because the GAT is meant for extracting relevant information according to inductive learning. In this case, the license plate number is the relevant information for GAT. This results in Adaptive Progressive Scalable Expansion based Graph Attention Network (APSEGAT), which outputs saliency for the input images. The saliency is fed to PSE algorithm for license plate number detection in the images. The flow of the proposed method can be seen in Fig. 2.

3.1. Adaptive Progressive Scalable Expansion based Graph Attention Network (APSEGAT)

Given an Image $I \in \mathbb{R}^{h\times w\times 3}$, we construct an FPN (Feature Pyramid Network) denoted by $\boldsymbol{F}$ consisting of a ResNet-101 feature extractor, which outputs feature maps denoted as: ($C2 \in \mathbb{R}^{h2\times w2\times c2}$, $C3 \in \mathbb{R}^{h3\times w3\times c3}$, $C4 \in \mathbb{R}^{h4\times w4\times c4}$, $C5 \in \mathbb{R}^{h5\times w5\times c5}$), where {h1, h2, h3, h4} represents the height, {w2, w3, w4, w5} represents width, and {c2, c3, c4, c5} represents the number of channels of the output feature maps. Consider a feature map is fed to a (1×1) convolution layer followed by batch normalization and rectified linear units (ReLU) activation function to reduce the number of channels to 256, be represented by $\boldsymbol{Tr}(\cdot)$, we calculate the following resulting feature maps as: $\mathbf{P_5} = \boldsymbol{Tr}(\mathbf{C_5})$; $\mathbf{P_4} = \boldsymbol{Tr}(\mathbf{C_4}) + \boldsymbol{Up}_{\times 2}(\boldsymbol{P_5})$; $\mathbf{P_3} = \boldsymbol{Tr}(\mathbf{C_3}) + \boldsymbol{Up}_{\times 2}(\boldsymbol{P_4})$; $\mathbf{P_2} = \boldsymbol{Tr}(\mathbf{C_2}) + \boldsymbol{Up}_{\times 2}(\boldsymbol{P_3})$; where $\mathcal{P}2 \in \mathbb{R}^{h2\times w2\times 256}$, $\mathcal{P}3 \in \mathbb{R}^{h3\times w3\times 256}$, $\mathcal{P}4 \in \mathbb{R}^{h4\times w4\times 256}$, and $\mathcal{P}5 \in \mathbb{R}^{h5\times w5\times 256}$. The proposed method concatenates the output of feature maps to generate a fused vector z as defined in Equation (1).

$$z = P_2 || Up_{\times 2}(P_3) || Up_{\times 4}(P_4) || Up_{\times 8}(P_5) \quad (1)$$

where "||" is concatenation operation and $\mathrm{Up}_{\times\alpha}(\cdot)$ upsamples the feature map α times.

The fused vector z is then fed to n (1×1) Convolution layers followed by Batch Normalization (BN), ReLU, and upsampling layer to produce n segmentation maps denoted as $S = \{S_1, S_2, \ldots, S_n\}$. The Progressive Scale Expansion algorithm (PSE) is applied on these n segmentation maps to detect the ground-truth region.

The obtained $\boldsymbol{C}` \in \{C4, C5\}$ from the above step PSENet is passed to a GAT layer. Since GAT layer expects a graph along with its adjacency matrix, the proposed method segments the grid-like structure $\boldsymbol{C}` \in \mathbb{R}^{h` \times w` \times c`}$ into multiple local feature maps $\boldsymbol{\psi}`_{\mathbf{k}}$, where each local feature is considered as a node of the graph. Here, every node is connected to each other representing some distant local features. Therefore, the adjacency matrix of this graph is a matrix of ones that allows information exchange of a local region with all distant patches.

Since the input image size varies, the height and width of its corresponding feature map $\boldsymbol{C}`$ also changes. To overcome this problem, segmentation of $\boldsymbol{C}`$ using a sliding window mechanism with a dynamically calculated stride and window size is proposed, which results in the same number of local patches where each patch is of different sizes. Each local patch is considered as a node in the graph. Therefore, the graph preserves the number of nodes as well as its mesh structure. We calculate the stride and window size for segmentation of $\boldsymbol{C}`$ using sliding window mechanism, dynamically at run-time, as defined in Equation (2).

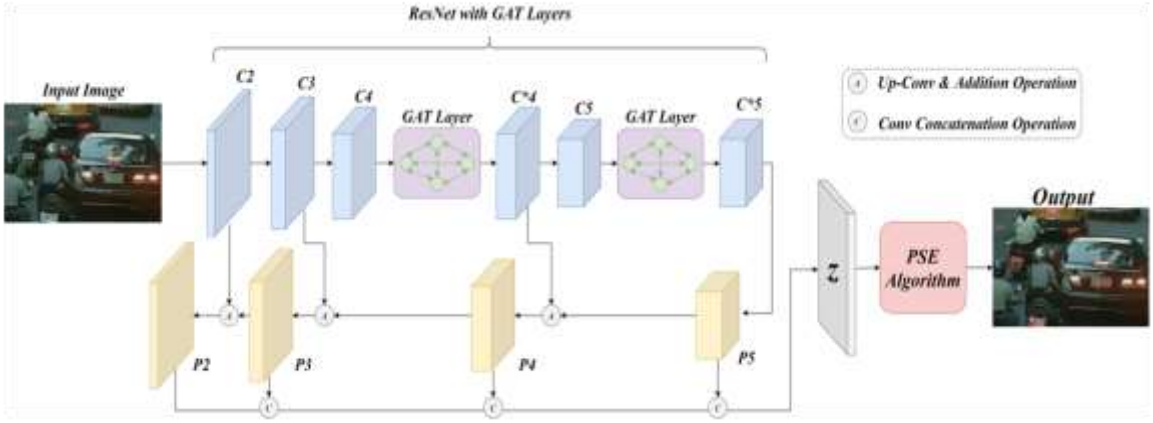


Fig. 3. Network architecture of the proposed APSEGAT

$$\begin{aligned} dh &= \lfloor h`/hh \rfloor \\ kh &= h` - dh * (hh - 1) \\ dw &= \lfloor w`/hw \rfloor \\ kw &= w` - dw * (hw - 1) \end{aligned} \tag{2}$$

where $\lfloor \times \rfloor$ is the floor function, dh and dw represent stride, kh and kw represent window size, hh and hw represent the number of desired nodes in our graph along the height and width, respectively. For all our experiment, the value of hh and hw is 10 which is determined experimentally. In other words, this process results in a graph of size (10×10) nodes where all nodes are connected to each other.

After patch extraction, we have: $\boldsymbol{\psi}` = \{\boldsymbol{\psi}`_1, \boldsymbol{\psi}`_2, \ldots, \boldsymbol{\psi}`_{100}\}$, where each $\boldsymbol{\psi}`_k$ have a dimension $\mathbb{R}^{0.1*h` \times 0.1*w` \times c`}$. This is followed by performing max pooling on each patch to give a resulting vector: $\boldsymbol{\psi}``_k \in \mathbb{R}^{1 \times 1 \times c`} \forall\ k \in [1, 100]$ and the corresponding max indices denoted by $\boldsymbol{\Gamma}$. It is noted that a node is also connected to itself, which is for self-attention mechanism. For every $C`$ we define a shared linear transformation weight matrix $\mathbf{W}` \in \mathbb{R}^{c` \times c`}$ to obtain sufficient expressive power that extracts higher-level features from $\boldsymbol{\psi}``_k$. This is followed by performing self-attention on the nodes using a shared attention mechanism, $a: \mathbb{R}^{c`} \times \mathbb{R}^{c`} \rightarrow \mathbb{R}$ as defined in Equation (3).

$$e_{xy} = a(\mathbf{W}`\mathbf{y}``_x, \mathbf{W}`\mathbf{y}``_y) \tag{3}$$

where, e_{xy} are attention coefficients that represents the extent of information exchange between node x and y.

We finally normalize the attention coefficients to make them comparable across different nodes using the softmax function as defined in Equation (4).

$$\alpha_{xy} = \frac{\exp(e_{xy})}{\sum_{k=1}^{100} \exp(e_{xk})} \tag{4}$$

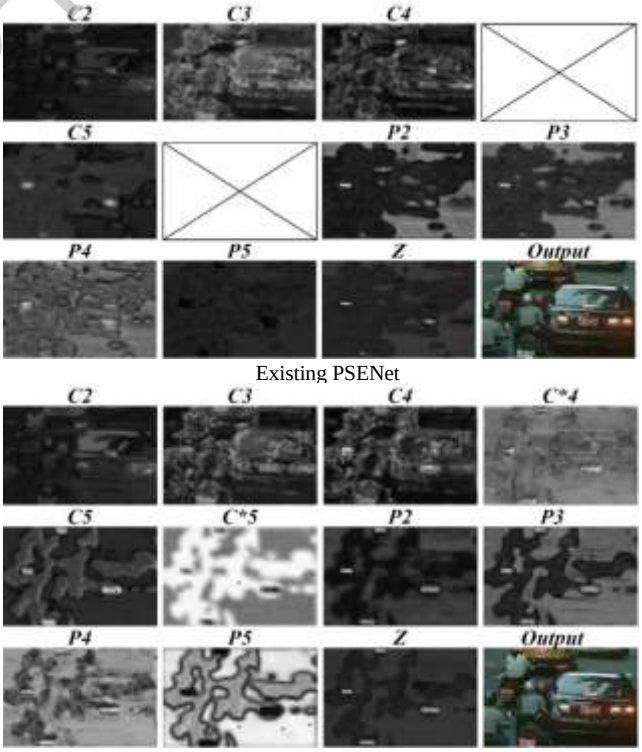
Further, a single layered feed-forward neural network is used as shared attention mechanism with the parameter of weight vector $a \in \mathbb{R}^{2c`}$ along with the LeakyReLU activation function, where the negative input slope is considered as 0.2. Therefore, attention coefficients are calculated as defined in Equation (5).

$$\alpha_{xy} = \frac{\exp(\text{LeakyReLU}(a^T[\mathbf{W}`\mathbf{y}``_x \,||\, \mathbf{W}`\mathbf{y}``_y]))}{\sum_{k=1}^{100} \exp(\text{LeakyReLU}(a^T[\mathbf{W}`\mathbf{y}``_x \,||\, \mathbf{W}`\mathbf{y}``_k]))} \tag{5}$$

The proposed method evaluates the feature representation corresponding to each node using a linear combination as defined in Equation (6).

$$y`^*_x = sigmoid(\textstyle\sum_{k=1}^{100} \alpha_{xk} W`y``_k) \tag{6}$$

Next, we perform unmax pooling operation on $y`^*_x$ using the max indices $\boldsymbol{\Gamma}$ to get: $y^*_x \in \mathbb{R}^{0.1h` \times 0.1w` \times c`}$. In order to add the knowledge of global context into $\psi`_k$, we perform channel-wise concatenation of $y`^*_x$ and $\psi`_k$ which is followed by a $(1 \times 1)$ convolution layer to reduce the number of channels to $c`$. Therefore, $C4$ from the PSENet is updated to $C*4$ followed by $C5$ to $C*5$ forming a hierarchical GAT structure as shown in Fig. 3 to have a complete architecture. The effect of the proposed APSEGAT including activation map of each layer can be seen in Fig. 4. Here we can see the existing PSENet does not sharpen key information in the images while the proposed Adaptive PSENet-GAT is able to sharpen the key information, like motorbike and car information in the images. This shows that the proposed method does not miss license plate information irrespective of challenges.



Existing PSENet

Proposed APSEGAT

Fig. 4. Activation maps from each layer of the existing PSENet and the Proposed APSEGAT

The evidence of detect accurate license plate number is illustrated in Fig. 5, where the standard deviation for the coefficients of the proposed model on different license plate numbers is computed. It is observed from Fig.5 that the variation among coefficients of each license plate number is almost similar.

This is the common property (the relationship) that is shared by license plate numbers across vehicles in the same image.

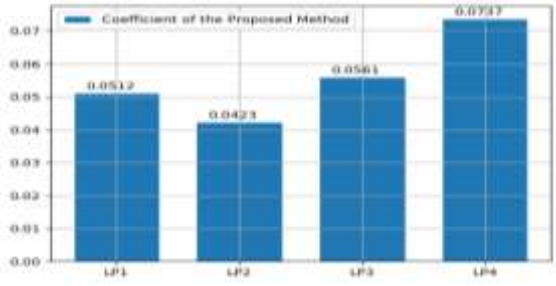


Fig.5. Standard deviation for the coefficients of the proposed model on different license plate numbers.

3.2. End-to-End Training Mechanism for License Plate Detection in Dense Vehicles Images

Consider the ground-truth segmentation maps corresponding to $\boldsymbol{S}$ discussed in the previous section be $\boldsymbol{G} = \{\boldsymbol{G_1}, \boldsymbol{G_2}, \dots, \boldsymbol{G_n}\}$, and the training mask be $\boldsymbol{M}$, we calculate the loss $\boldsymbol{L}$ similar to [5] as follows:

$$
\begin{aligned}
L_c &= 1 - Dice(S_n \cdot M, G_n \cdot M) \\
L_s &= 1 - \frac{\sum_{i=1}^{n-1} Dice(S_i \cdot W, G_i \cdot W)}{n-1} \\
L &= lL_c + (1-l)L_s \\
where,\ W_{x,y} &= \begin{cases} 1, & if\ S_{n,x,y} \geq 0.5; \\ 0, & otherwise \end{cases} \\
Dice(S_i,\ G_i) &= \frac{2\sum_{x,y} S_{i,x,y} * G_{i,x,y}}{\sum_{x,y} S_{i,x,y}^2 + \sum_{x,y} G_{i,x,y}^2} \\
\theta^F &= \theta^F - \varepsilon \Delta L
\end{aligned}
\tag{7}
$$

where $\boldsymbol{\theta^F}$ are the model parameters and $\boldsymbol{\varepsilon}$ is the learning rate.

During training, the input images are normalized to standard size of 640×640 by padding zeros. Experiments are conducted on a computer with Intel i5-8600 CPU and NVIDIA GeForce GTX 1070 Ti. The entire model is implemented using PyTorch. We use Adam Optimizer [22] with batch size3, and learning rate 0.0001where weight decay is 0 .0005 for 300 epochs. The value of $l$in Equation (7) is chosen as 0.7 following our experiments on predefined samples. The number of segmentation maps are predicted by the network is 6 as presented in the previous section.

4. Experimental Results

For evaluating the proposed method for detecting license plate numbers in dense vehicles images captured in crowded street scenes, we create our own dataset because there is no standard dataset available publicly. Our dataset named as (AMLPR) includes images of different types of cars and motorbikes as shown in sample images in Fig. 6(a). It is noted from Fig. 6(a) that due to dense vehicles including different motorbikes and cars in a single image, the license plate number loses quality and there are high chances of occlusion. At the same time, we can also expect degradations and poor quality of images due to defocus of the camera on the many vehicles. Therefore, our dataset is challenging for license plate number detection. Our dataset consists of 1501 images with 4986 license plate number instances. More details of our dataset are reported in Table 1, where LP denotes License Plate and 1, 2, 3, 4, 5 and greater than 5 indicate the number of license plate instances in a single image. Note that the size of the input images is converted to standard size of 640×640 dimension for all the experiments in this work.

Table 1. Detail information of our dataset.

Total Images	Total Number of instances	Top row: number of license plate number in each image. Bottom row: number of images.					
		1 LP	2 LP	3 LP	4 LP	5 LP	>5 LP
1501	4986	113	225	562	340	192	66

To test the ability of the proposed method on the images that containing single vehicle in single image, we consider two benchmark datasets: (i) UFPR-ALPR and (ii) UCSD. In UFPR-ALPR dataset most of the images contain single vehicle in single image with different background and car types. This dataset provides 4500 images for experimentation [16]. In addition, most of the images are captured in regular traffic and urban environment. As a result, one can expect complex background images affected by distance variation between camera and target as shown sample images in Fig. 6(b). UCSD dataset [6] provides images affected by different orientations and high-quality images as shown in Fig. 6(c), where we can see high quality images containing different car types. This dataset provides 291 images. The reason to consider these two benchmark datasets is that they provide different nature, characteristics and complexities. In summary, 6292 images are considered for experimentation in this work. During training and testing, we consider 75% of the samples for training and 25% of the samples for testing. The same set up is used for all the experiments on all the datasets.

(a) Our AMLPR Dataset

(b) UFPR-ALPR [16] Dataset

(c) UCSD Dataset [6]

Fig.6. Sample images of our and different benchmark datasets.

To show the effectiveness of the proposed method, we implemented the YOLO [9] and PSENet [5] methods for comparative study in this work. It is noted from the existing methods [8, 9, 10] in the literature for license plate number detection that the recent methods use YOLO architecture as the main base for achieving better results on different conditions and situations. Therefore, we use the method [9] that employs YOLO architecture for license plate detection to compare with the proposed method on all the three datasets. In addition, this is also to show that the method is not adequate to addresses the challenges of the images containing dense vehicles in crowd. Similarly, the PSENet architecture is the state-of-the-art network for text detection in natural scene images, which is proposed to address several challenges of natural scene text detection, such as complex background, different orientations, font size and contrast variation similar to challenges of license plate number detection. In addition, the proposed method adapts the same base of PSENet architecture with modifications for license plate number

detection in this work. To judge the effectiveness of APSEGAT over the existing PSENet, we compare the performance of the existing PSENet with the performance of the proposed method on all the three datasets. Further, the methods which use PSENet for natural scene text detection may not work well for the image of dense vehicles.

To measure the performance of the proposed and the existing methods, we use the standard measures, namely, Recall (R), Precision (P) and F-Measure (F) and we follow the instructions and definition as mentioned in the method of [6] for all the experimentations.

4.1. Experiments on our Dense Vehicles Dataset

Qualitative results of the proposed method for the sample images of our dataset are shown in Fig. 7, where we can see that although, the license plate numbers miss some information due to partial occlusion, the proposed method detects all of them successfully. This is the key advantage of the proposed method in addition to detecting multiple license plate numbers of different vehicles in a single image. It is also observed that the license plate number of motorbikes are degraded, have small font and not visible properly. The proposed method detects all the license plate number accurately. Therefore, we can conclude that the proposed method is independent of the number of vehicles and type of vehicles.

Fig. 7. Qualitaitve results of the proposed method on our dataset.

Quantitative results of the proposed and existing architectures are reported in Table 2, where one can note that the proposed method is the best at F-measure compared to other methods. However, the PSENet achieves the best Precision while YOLO based method [9] achieves the best Recall compared to the proposed method. This shows that the method [9] is good for detecting license plate numbers, but it gives a greater number of false positives for the images with many vehicles of different types. The reason is that the method [9] is developed for license plate number detection in the images having single vehicle. In case of PSENet, high precision indicates that the PSENet is good for license plate detection with a smaller number of false positives. In addition, due to occlusion and small font, the PSENet misses license plate numbers in the images and hence recall is lower than the method [9] and the proposed method. The reason is that the PSENet is developed for natural scene text detection. Overall, the proposed method achieves the highest F-measure compared to the method [9] and PSENet. This shows that the proposed Adaptive Progressive Scalable based Graph Attention Network (APSEGAT) has ability to cope with the challenges of the images of crowded street scenes.

4.2. Experiments on Benchmark Dataset

Sample qualitative results of the proposed method for UFPR-ALPR and UCSD datasets are shown in Fig. 8(a) and Fig. 8(b), respectively. It is noted from Fig. 8 that the proposed method well detects license plate number for different vehicles with a different background. Note that in these datasets, most of the images contain a single type vehicle. The results in Fig.8 show that the proposed method is capable of achieving good results for the images containing a single type vehicle with cluttered backgrounds.

Quantitative results of the proposed and existing methods are reported for UFPR-ALPR and UCSD datasets in Table 2. It is observed from Table 2 that the proposed method achieves the best F-measure compared to the existing methods. The method [9] achieves the highest recall for both the dataset compared to other existing methods and the proposed method. This is obvious because the method [9] is designed for license plate number detection in the images having a single vehicle while PSENet is designed for natural scene text detection and hence it reports poor results for these datasets. Overall, when we look at F-measure of the method [9] and PSENet, the PSENet is better than the method [9] for UFPR-ALPR while lower than the method [9] for UCSD datasets. This show that the method [9] is good for high-quality images containing one type of vehicle compared to the images of crowded scenes, which suffer from poor quality caused by occlusion and varying degree of focus for the different vehicles.

When we compare the results of our, UFPR-ALPR and UCSD datasets, the performance of the methods improves slightly for UFPR-ALPR and UCSD datasets compared to our dataset. This confirms that our dataset is more challenging for this problem than the standard datasets.

(a) UFPR-ALPR [16]

(b) UCSD [6]

Fig. 8. Qualitative result of the proposed methhod on benhmark datasets.

Table 2. Comparative performances of the proposed and the existing methods on our and benchmark datasets.

Methods	AMLPR			UFPR-ALPR			UCSD		
Measures	P	R	F	P	R	F	P	R	F
YOLO [9]	0.81	**0.92**	0.86	0.75	**0.98**	0.85	0.95	**1.0**	0.97
PSENet [5]	**0.95**	0.79	0.86	0.98	0.92	0.95	0.95	0.92	0.03
Proposed	0.92	0.86	**0.90**	**0.99**	0.95	**0.97**	**0.97**	**1.0**	**0.98**

When we compare the average processing time of the proposed and existing models for license plate number detection, the YOLO consumes lowest processing time, which is around 0.03 seconds for processing each image. The PSENet and the proposed model consume around 1.00 second for processing each

image. This is acceptable because the YOLO is designed for fast processing while PSENet and the proposed model are not. Since the PSENet involves multi-scaling operation in the progressive scale expansion (PSE) algorithm to accumulate the information at the pixel level which requires more processing time, it is obvious that our model consumes time slightly higher than YOLO. However, the main objective of the proposed work is to find a solution to the complex problem of multiple license plate number detection in crowded scenes. Therefore, the proposed work does not focus on time efficiency. In addition, due to higher-end GPU systems, processing time does not affect for the overall performance of the proposed method unless it is in real-time environment where efficiency is vital. Furthermore, the processing time depends on several factors such as the use of platform for coding, system configuration and programming language. Hence, we can conclude that the marginal processing time difference between the proposed model and the YOLO can be ignored.

Fig. 9. Unsuccessful results of the proposed method.

Sometimes, when images are affected by heavy light illumination and lose visibility as shown in Fig. 9, the proposed method does not perform well. The reason is that APSEGAT loses ability to differentiate background and license plate numbers at pixel level. For these cases, it is necessary to enhance fine details to improve results, which is beyond the scope of the work.

5. Conclusion and Future Work

We have proposed a new method for detecting license plate number in crowded street scenes. To achieve our goal, the method adopts PSENet for extracting unique information which represents license plates. At the same time, to enhance the ability to cope with the challenges caused by partial occlusion and varying degree of focus for different vehicles, the method integrates Graph Attention Network (GAT) with Adaptive PSENet. We have tested it on our dataset and two benchmark datasets. The results show that the method outperforms the existing methods in terms of F-measure for all the three datasets. However, when image suffers from loss of visibility and severe glare, the method does not perform well. Therefore, we plan to integrate new networks for enhancement to improve the results in the future.

Acknowledgments

The work has received the Faculty Grant: GPF014D-2019, University of Malaya, and the Natural Science Foundation of China associated with Grant 61672273. The authors of this paper acknowledge Anirban Saha and Ananta Kumar Ghosh, Indian Statistical Institute, Kolkata, India for their help in annotating the dataset for experimentation.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. R. Asif, C. Qi, T. Wang, M. S. Fareed and S. A. Raza, “License plate detection for multi-national vehicles: An illumination invariant approach in multi-lane environment”, Computers and Electrical Engineering, 78, pp 132-147, 2019.

[2] Z. Slemi, M. B. Halima, U. Pal and M. A. Alimi, “DELP-DAR system for license plate detection and recognition”, Pattern Recognition Letters, 129, pp 213-223, 2020.

[3] N. F. Gazcon, C. I. Chesenevar and S. M. Castro, “Automatic vehicle identification for Argentinean license plates using intelligent template matching”, Pattern Recognition Letters, 33, pp 1066-1074, 2012.

[4] M. F. Sadique and S. M. R .Haque, “A comparative study of license plate detection and recognition techniques”, In Proc. ICCT, 2019.

[5] W. Wang, E. Xie, X. Li, W. Hou, T. Lu, G. Yu, S. Shao, “Shape Robust Text Detection With Progressive Scale Expansion Network,”, in Proc. CVPR, pp 9328-93337, 2019.

[6] P. Shivakumara, S. Roy, H. A .Jalab, R.W. Ibrahim, U. Pal, T. Lu, V. Khare and 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.

[7] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement”, arXiv, 2018.

[8] Hendry and R. C. Chen, “Automatic license plate recognition via sliding window darknet-YOLO deep learning”, Image and Vision Computing, 87, pp 47-56, 2019.

[9] Y. Kesentini, M. D. Besbes, S. Ammar and A Chabbouh, “A two-sage deep neural network for multi-norm license plate detection and recognition”, Expert Systems with Applications, 136, pp 159-170, 2019.

[10] W. Min, X. Li, Q. Wang, Q. Zeng and Y. Liao, “New approach to vehicle license plate location based on new model YOLO-L and plate pre-identification”, IET-Image Processing, 13, pp 1041-1049, 2019.

[11] 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.

[12] C. Liu and F. Chang, “Hybrid cascade structure for license plate detection in large visual surveillance scenes”, IEEE Trans. ITS, 20, pp 2122-2135, 2019.

[13] M. M. Moreno, I. G. Diaz and F. D. D. Maria, “Efficient scale-adaptive license plate detection system”, IEEE Trans. ITS, 20, pp 2109-2121, 2019.

[14] N. Omar, A. Sengur and S. G. S. Al-Ali, “Cascaded deep learning based efficient approach for license plate detection and recognition”, Expert Systems with Applications, 149, 2020.

[15] M. S. Al-Shemarry, Y. Li and S. Abdulla, “An efficient texture descriptor for the detection of license plate from vehicle images in difficult conditions”, IEEE Trans. ITS, pp 553-564, 2020.

[16] S. M. Silva and C. R. Jung, “Real-time license plate detection and recognition using deep convolutional neural networks”, Journal of Visual Communication and Image Representation, 2020.

[17] R. Bhargav and P. Deshpande, “Locating multiple license plates using scale, rotation and color-independent clustering and filtering techniques”, IET-Image Processing, 13, pp 2235-2345, 2019.

[18] Y. Luo, Y. Li, S. Huang and F. Han, “Multiple Chinese vehicle license plate localization in complex scenes”, In Proc. ICIVC, pp 745-749, 2018.

[19] A. Menon and B. Omman, “Detection and recognition of multiple license plate from still images”, In Proc. ICCSDET, 2018.

[20] Y. Zhou, M. N. Lv, Z. Q. Ling and D.L. Li, “Multiple license plate location algorithm in complex scene”, In Proc. IUCC, pp 456-461, 2019.

[21] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, Y. Bengio, “Graph Attention Networks”, In Proc. ICLR, pp 1-12, 2018.

[22] D. P. Kingma, and J. Ba, “Adam: A Method for Stochastic Optimization”, In Proc. ICLR, pp 1-15, 2015.

Graphical Abstract (Optional)

Graph Attention Network for Detecting License Plates in Crowded Street Scenes

Leave this area blank for abstract info.

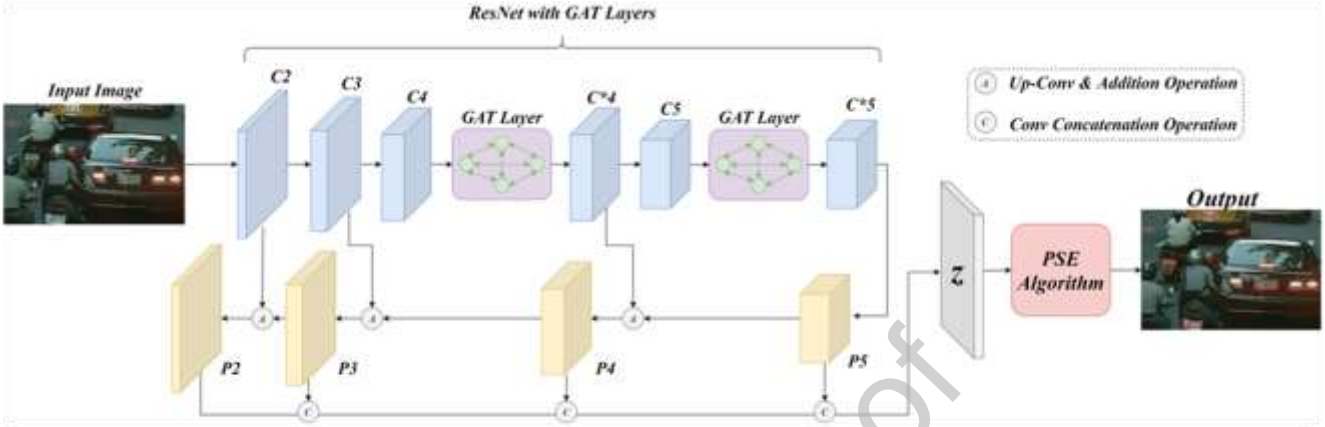


Fig. 3. Network architecture demonstrating the Proposed Method.