License Plate Detection via Information Maximization

Younkwan Lee, *Student Member, IEEE*, Jihyo Jeon, *Student Member, IEEE*, Yeongmin Ko, *Student Member, IEEE*, Moongu Jeon, *Senior Member, IEEE*, and Witold Pedrycz, *Life Fellow, IEEE*

***Abstract*— License plate (LP) detection in the wild remains challenging due to the diversity of environmental conditions. Nevertheless, prior solutions have focused on controlled environments, such as when LP images frequently emerge as from an approximately frontal viewpoint and without scene text which might be mistaken for an LP. However, even for state-of-the-art object detectors, their detection performance is not satisfactory for real-world environments, suffering from various types of degradation. To solve these problems, we propose a novel end-to-end framework for robust LP detection, designed for such challenging settings. Our contribution is threefold: (1) A novel information-theoretic learning that takes advantage of a shared encoder, an LP detector and a scene text detector (excluding LP) simultaneously; (2) Localization refinement for generalizing the bounding box regression network to complement ambiguous detection results; (3) a large-scale, comprehensive dataset, LPST-110K, representing real-world unconstrained scenes including scene text annotations. Computational tests show that the proposed model outperforms other state-of-the-art methods on a variety of challenging datasets.**

***Index Terms*— License plate detection, deep learning, information theory, multi-task learning, intelligent traffic surveillance.**

I. Introduction

OBJECT detection research has attracted great interest in recent years, with models being applied widely in many traffic-related applications [1]–[6]. A variety of methods have demonstrated high accuracy in detecting license plates (LP) under controlled settings.

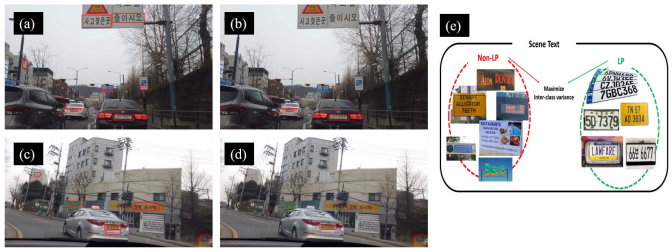


Fig. 1. **Detection in wild scenes and an illustration of license plate (LP) vs non-LP class.** A typical image in our LPST-110K, showing unconstrained settings. The first column (a, c) is detection results for the state-of-the-art RetinaNet [4]. The second column (b, d) shows the our results, indicating fewer detection errors and better regression. The last column (e) is an illustration of scene text relation.

While existing detectors successfully applied to the LP detection problem, many key challenges still remain in *unconstrained wild scenarios*. For example, real-world LP detection causes the following problems: modifications of prior settings to adapt to wild, incorrect detection results, ambiguity in classifying objects associated with scene text, low-quality visual data, uneven lighting, motion blur, and others. However, such scenarios are becoming increasingly common and gaining significant popularity in a variety of applications, including civil security, crowd analytics, law enforcement, and street view images. Despite being the most common scenario, LP benchmarks still do not consider real-world cases, and therefore many problems are not adequately addressed. As a result, state-of-the-art detectors struggle with these images.

To clearly ascertain what makes LP detection difficult, some common cases in the wild must be considered where LP and scene text appear at the same time as multiple instances (see Figure 1). Based on this basic observation, we identify two major drawbacks in two aspects. First, LP and the scene text (not LP) are not correctly distinguished, which in return may cause false detection of each other. In fact, the LP is a child class that belongs to the scene text, so they must be distinguished and there must be enough variability to distinguish class categories. The existing LP benchmarks, however, did not include scene text in the sample, nor were they explicitly addressed in learning and evaluation. Secondly,

Manuscript received 4 August 2020; revised 4 July 2021 and 15 October 2021; accepted 7 December 2021. Date of publication 24 December 2021; date of current version 12 September 2022. This work was supported by the Institute of Information and Communications Technology Planning and Evaluation (IITP) funded by the Government of Korea (Ministry of Science and ICT, Development of Global Multi-Target Tracking and Event Prediction Techniques Based on Real-Time Large-Scale Video Analysis) under Grant 2014-3-00077. The work of Witold Pedrycz was supported in part by the National Research Foundation of Korea (NRF) funded by the Korea Government (Ministry of Science and ICT) under Grant 2019R1A2C2087489, in part by the Korea Creative Content Agency (KOCCA) funded by the Government of Korea (Ministry of Culture, Sports and Tourism) under Grant R2020070004, and in part by the National Natural Science Foundation of China under Grant 62076182. The Associate Editor for this article was H. Jula. *(Corresponding author: Moongu Jeon.)*

Younkwan Lee, Jihyo Jeon, Yeongmin Ko, and Moongu Jeon are with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, South Korea (e-mail: brightyoun@gist.ac.kr; jihyo@gist.ac.kr; koyeongmin@gist.ac.kr; mgjeon@gist.ac.kr).

Witold Pedrycz is with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB T6R 2V4, Canada, also with the Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah 21589, Saudi Arabia, and also with the Systems Research Institute, Polish Academy of Sciences, Warsaw 01-447, Poland (e-mail: wpedrycz@ualberta.ca).

Digital Object Identifier 10.1109/TITS.2021.3135015

the detected bounding box does not contain all the characters in the LP. Basically, LP detection is necessarily linked to continuing tasks related to recognition or de-identification; therefore, sophisticated localization is essential for identifying information. Yet, for detailed extra tasks, it is still challenging to localize enough information contained in LPs. Interestingly, as shown in Fig.1 (a, c), the state-of-the-art detector exhibits prominent negative results for scenarios in the wild.

A well-designed LP detection framework should tackle the problems above (see in Figure 1(b, d)). In this paper, we propose an end-to-end framework which is composed of a single shared feature encoder and two parallel detection branches. The single shared encoder learns a global feature across all detection tasks (LP and non-LP respectively). More specifically, due to non-LP objects (scene text but not LP), our framework is divided into 1) LP detection network and 2) non-LP detection network. Different from traditional LP detection models, we explicitly prevent learning of non-LP objects. To this end, we bring a novel *information-theoretic loss* to minimize mutual information between the embedding feature and non-LP distribution that interferes with LP detection. Prior to the unlearning of non-LP distribution, we hypothesize that the existence of non-LP is known and that the relevant meta-data, such as additional labels corresponding to the semantics of the non-LP instances are accessible. In this scenario, the discrimination problem between LP and non-LP based on mutual information can be formulated in terms of an adversarial problem. One network has been trained to detect the non-LP instances. Instead, the other network has been trained to detect only LP instances, which is the ultimate goal of the overall architecture, while maximizing the discrimination between LP and non-LP based on mutual information. Therefore, we adopt an adversarial training strategy, which is achieved by minimizing mutual information while estimating optimal LP detection independence. Furthermore, we propose a *localization refinement* module with a sharing block. This module provides valuable information on the quality of bounding box regression for sophisticated localization.

To summarize, this paper makes the following novel contributions:

- **A novel information-theoretic loss for LP detection.** We propose a new framework that is discriminative to detect LP even in unconstrained scenes. We note that our approach to calculating mutual information could likely exclude non-LP, resulting in high accuracy (Sec. III.C).
- **Localization refinement module.** We generalize the bounding box regression network to complement ambiguous detection results. As far as we know, there has been no other previous work to utilize regression networks for refinement of localization (Sec. III.D).
- **A novel LP detection dataset.** We collect a new large-scale dataset, **LPST-110K**, containing images captured from unconstrained scenes. To the best of our knowledge, LPST-110K is the first dataset to address LP and scene text simultaneously for LP detection. By evaluating state-of-the-art detection models on LPST-110K, we demonstrate the accuracy improvement of our proposed model compared with other approaches (Sec. IV).

II. Related Works

In this section, we review the deep learning algorithms in *intelligent transportation systems* (ITS) and the LP detection methods related to our methods. The deep learning in ITS, the license plate detection and license plate detection benchmarks are included in this section.

A. Deep Learning in ITS

In recent years, deep learning algorithms have achieved impressive results in computer vision [7]–[10]. In many modern transportation systems, deep learning has begun to play a critical role as a means to acquire more robust recognition or surveillance, by learning from existing task-specific benchmarks. It is performed to solve more complex traffic conditions by designing a non-linear model based on a data-driven paradigm with existing benchmarks. Many traditional problems such as road detection [11], [12], street scene labeling/recognition [13], [14], crowd counting [15], [16], traffic flow estimation [17], [18], or license plate detection [19], [20] and recognition [21]–[23] can be investigated to utilize these techniques. Specifically, depending on the existing benchmarks and detection algorithms, robust license plate detection can help take to help guide a more comprehensive understanding and control of traffic conditions. While researchers have utilized limited benchmarks and universal detection algorithms, we have found that conventional algorithms are not always the solution in every situation. Developing a more robust solution is a non-trivial task, but is required to outperform current capabilities. We therefore investigate what efforts and trials have been made in prior works for license plate detection algorithms and benchmarks in the following subsections.

B. License Plate Detection

Early works have devoted much effort to improving LP detection performance based on the framework of image binarization model [24], [25], segmentation model [26], edge-based model [27], and region-based model [28]. In this way, several approaches have remarkably shown the use of different hierarchical schemes for detecting a vehicle region as part of extracting the LP region. Nevertheless, these methods cannot perform well on complex backgrounds and in unconstrained settings.

More recently, as Deep Convolutional Neural Networks (DCNN) [29], [30] have shown good classification performance, researchers have begun to deal with some complicated situations. Particularly, as deep feature-based object detectors [6], [31] have been developed, many studies have started to detect LP under difficult situations. Prior knowledge-based methods based on vehicle detection [19], [32]–[38] have greatly reduced false positives despite background clutter. Data-driven methods [35], [39]–[42] have been used to increase the detection accuracy by exploiting useful deep representations with the augmentation transforms. Specifically, [20], [35], [41] may be the most similar to ours, because they also focus on unconstrained environments. However, these studies still do not consider the existence of non-LP, thus have not reached a wide diffusion. Our work is distinguishable in that we try to address the non-LP instance in unconstrained

TABLE I

KEY PROPERTIES FOR LP DETECTION BENCHMARKS. #IMAGE.: NUMBER OF IMAGES. W. ST.: THE PRESENCE OF SCENE TEXT IN THE IMAGE. #INSTANCE.: NUMBER OF INSTANCES WITH BOUNDING BOX ANNOTATION. #LP/IMAGE.: AVERAGE LPS PER IMAGE. #ST/IMAGE.: AVERAGE SCENE TEXT (LP AND NON-LP) INSTANCES PER IMAGE. VARIATIONS IN TILT DEGREES.: GREAT HORIZONTAL TILT DEGREE (15°∼45°) AND VERTICAL TILT DEGREE (15°∼45°). VARIATIONS IN DISTANCE.: THE DISTANCE FROM THE LP TO THE CAMERA LOCATION IS RELATIVELY DIVERSE. VARIATIONS IN BLUR.: BLURRY IMAGE DUE TO MOTION BLUR AND HAND JITTER WHILE CAPTURING IMAGES

Name (Year)	#Image	w. ST.	#Instance	#LP/Image.	#ST/Image.	Variations in tilt degrees.	Variations in distance.	Variations in blur.
AOLP (2013) [43]	2,049	✗	2,049	1	1	✓	✗	✓
SSIG (2015) [44]	2,000	✗	8,683	4.34	4.34	✗	✓	✗
PKU (2016) [45]	3,977	✗	4,389	1.10	1.10	✗	✗	✗
UFPR (2018) [33]	4,500	✗	4,500	1	1	✗	✓	✓
CD-HARD (2018) [35]	102	✗	102	1	1	✓	✓	✗
CCPD (2018) [41]	**250K**	✗	**250K**	1	1	✓	✓	✓
Ours LPST-110K	9,795	✓	110K	**5.21**	**11**	✓	✓	✓

cases. Moreover, our experiments show that our completed method improves LP detection performance in the real-world scenarios.

C. License Plate Detection Benchmarks

Many benchmarks for LP detection were designed for training and testing simultaneously and a few surveys are shown in Table I. Representative LP detection datasets include AOLP [43], SSIG [44], PKU [45], CD-HARD [35], UFPR [33] and CCPD [41]. Surprisingly, none of these provide scene-text annotations, even though they are the main cause of the erroneous detection.

As evident in Table I, our new LPST-110K dataset, described in Sec. IV, provides all text annotations that exist in the image that have not been attempted in any datasets. Moreover, our datasets, which focused on rough scenes in uncontrolled environments, were challenging and particularly related to motion blur, uneven lighting, large slope angle and low resolution. The exceptions are UFPR [43] and CCPD [41], which consist of many of the aforementioned non-constraining conditions. In particular, the CCPD [41] provides a huge number of samples that cannot be compared with other benchmarks. Despite this fact, these images provide only one to three samples per image, but LPST-110K provides as few as three to as many as 20 LP annotations per image. More importantly, classification of LPs and non-LP texts on the LPST-110K gets confused between each other, making them a challenge for detection. To our knowledge, LPST-110K is the first dataset to provide text annotations as well as enormous numbers of instances (LP and non-LP) in an image, even collected from unconstrained scenes.

III. PROPOSED METHODOLOGY FOR LICENSE PLATE DETECTION

In this section, we first introduce the problem settings, which will be discussed in Section III-A. We then present the license plate detection architecture used in our experiments in Section III-B. In addition, we formulate the loss functions for each part of the whole architecture in detail (Section III-C-III-D) and define the overall training procedure, described in Section III-E. Finally, we illustrate how to perform the inference for the proposed model in Section III-F.

A. Problem Settings

In order to make the descriptions clear, we introduce several notation prior to the introduction of the overall idea of the study. Unless noted otherwise, all notations refer to the following terms. All the symbols and notation used in this paper are summarized in Table II. As shown in Fig 1, our goal is to detect LP from each image example $x \in \mathcal{X}$, where $\mathcal{X}$ denotes an input space for images. Then, the input image x contains an LP $y(x) \in \mathcal{Y}$ and a non-LP scene text $n(x) \in \mathcal{N}$ classification and 4-tuples bounding box coordinate labels. Let X and Y be two random variables. In this paper, we consider X and Y include the value of x and $y(x)$ respectively. We also represent $\mathcal{N}$ and $\mathcal{Y}$ as an non-LP class that interferes with LP detection and a LP class respectively. In addition, we define a latent function $n : \mathcal{X} \rightarrow \mathcal{N}$, where $n(x)$ denotes the target non-LP instance of x.

As already mentioned, our proposed network takes the input image x and outputs both LP detection $y(x)$ and non-LP detection $n(x)$ results simultaneously. Thus the input image x is fed into the encoder (ResNet + FPN) for feature extraction $f : \mathcal{X} \rightarrow \mathbb{R}^K$, where K is the number of the features extracted by f, parametrized as θ_f. Additionally, we replace the original RPN structure with two parallel RPN structures: RPN for LP $g : \mathbb{R}^K \rightarrow \mathcal{Y}$ and RPN for non-LP $h : \mathbb{R}^K \rightarrow \mathcal{N}$. The parameters of each network are denoted as $\theta_g \in [\theta_{gloc}, \theta_{gcls}]$ and $\theta_h \in [\theta_{hloc}, \theta_{hcls}]$, assuming the regression and classification sub-network parameters, respectively.

B. Architecture Design

As discussed in Section 1, we propose to utilize information-theoretic learning to improve the performance of LP detection, which aims to construct rich feature representations for complex and challenging scenes. As shown in Fig 2, our overall architecture is divided into three parts: 1) a backbone network f, 2) a LP detection sub-network g, and 3) a non-LP detection sub-network h. Existing two-stage detectors like Faster RCNN consist only of f and g, but our method additionally utilizes h to further maximize the discrimination between LP and non-LP in feature representation learning. Specifically, we include a localization refinement module (LRM) while learning g and h. It is worth mentioning that proposed architecture provides the complementary information to minimize mutual information

TABLE II
NOTATION USED IN THIS PAPER

Symbol	Meaning
x	The input image
$\mathcal{X}$	The input space for input images
$y(x)$	The license plate (LP) samples
$\mathcal{Y}$	The LP space for LP samples
$n(x)$	The non-LP scene text samples
$\mathcal{N}$	The non-LP scene text space for non-LP samples
X	The random variable for x
Y	The random variable for $y(x)$
f and θ_f	The backbone network and its network parameters
g and θ_g	The LP detection sub-network and its network parameters
h and θ_h	The non-LP detection sub-network and its network parameters
$\mathcal{S}$ and $l_{\mathcal{S}}$	The sharing block and its output
$\mathcal{I}(\cdot;\cdot)$	The mutual information between two random variables
$H(\cdot)$	The marginal entropy function
$H(\cdot\|\cdot)$	The conditional entropy function
$D_{KL}(\|\|)$	The KL divergence between two distributions
N and K	Total number of input images and the dimension number of the features

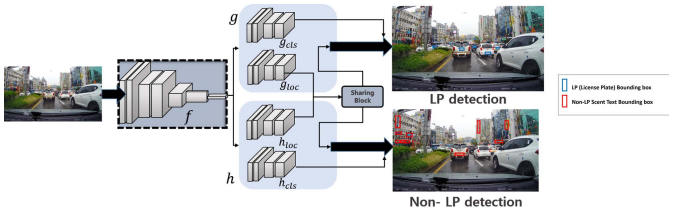


Fig. 2. **Overall architecture of LP detection network.** The network f is constructed with ResNet-50 and FPN.

between the embedding feature and non-LP distribution and boost LP-specific detection performance.

The input of our proposed architecture is the image x, the output is the LP and non-LP detection results for training and the only LP detection results for inference. A standard deep learning-based detection network is designed, motivated by [4], [31], [46]. First, the backbone network of ResNet-50 [47] is established by building FPN [46] with three upscaling-layers for feature extraction as an encoder f. Subsequently, our task-specific detection networks, well-known RPN [31], includes *two parallel structures* (*i.e.* one for LP g and the other for non-LP h), which provide two fully convolutional sub-networks. These sub-networks in RPN structures are attached to each feature map of the encoder network in parallel to each other.

The first is a *regression sub-network* which performs a bounding box regression for sophisticated localization around the object in the image using the encoder's output $f(x)$, represented as the x and y-axes coordinates in the upper-left corner and the x and y-axes coordinates in the lower-right corner of the rectangle. Secondly, *classification sub-network* produces a class-specific confidence score C_i, i denotes the number of classes including the background (assuming multi-class cases). Therefore, each anchor box has i numbers indicating the class probabilities.

C. Mutual Information Maximization via Adversarial Loss

In *constrained scenes*, one-class object detection task with only an LP class can improve the precision and localization accuracy with small false-positive rates and high-IoU scores simultaneously. In *unconstrained images*, however, there are scene-texts that look like LPs and arbitrarily shaped LPs. Thus, such phenomenon has produced unsatisfactory results in terms of LP detection performance. Ideally, LP-discriminative features should explicitly ignore non-LP related features inside the learned network. Therefore, for maximizing inter-class variance, the objective is to perfectly remove the following characteristics from detection network:

$$\mathcal{I}(Y; n(X)) \not\approx 0, \tag{1}$$

where $\mathcal{I}(\cdot)$ denotes the mutual information between two random variables. To handle this problem, our ultimate goal is to learn the network with the following characteristics:

$$\mathcal{I}(g(f(X)); n(X)) \approx 0. \tag{2}$$

We decide to add the mutual information term to the objective function for training networks. To be specific, during the training process, we should explicitly define a classification stage for non-LP, which aims to confuse non-LP data distribution from the extracted features.

We hope the LP-specific detector is trained to maximize to inter-class variations related to non-LP images. A good LP detector would, therefore, have characteristics that are close to the characteristics which are irrelevant for all non-LP visual representation, especially scene-text without LP. Therefore, we replace $g(f(X))$ with $f(X)$ because g, the RPN network that determines detection output, receives $f(X)$ as its input. This means that if the entire network recognizes $n(X)$, which is non-LP information, as disrupted information for LP detection, it already has that property from $f(X)$ extracted from the input image X. In this case, we derive the following objective function:

$$\min_{\theta_f, \theta_g} \mathcal{L}_{lp}(\theta_f, \theta_g) + \lambda_{obj}\mathcal{I}(f(X); n(X)), \tag{3}$$

where $\mathcal{L}_{lp}$ is the standard detection loss [4], [31], [46] including Euclidean loss for regression $\mathcal{L}_{gloc}$ and cross-entropy loss for classification $\mathcal{L}_{gcls}$. λ_{obj} is a trade-off hyper-parameter to control the relative importance of the two terms.

In information theory, the mutual information term in Eq. (3) can be explicitly expressed as follows:

$$\begin{aligned}\mathcal{I}(f(X); n(X)) &= H(f(X)) - H(f(X)|n(X)) \\ &= H(n(X)) - H(n(X)|f(X)),\end{aligned} \tag{4}$$

where $H(\cdot)$ and $H(\cdot|\cdot)$ are the marginal and conditional entropy, respectively. Here, the marginal entropy $H(n(X))$ can be eliminated from the objective function because it is a constant that is completely independent of θ_f and θ_g during the optimization process. However, the entropy term in Eq. (4) can be changed to the problem of calculating the posterior distribution. To be specific, we can instead calculate the negative conditional entropy $-H(n(X)|f(X))$ with the posterior $P(n(X)|f(X))$ explicitly. However, the posterior distribution in objective function is still intractable. We can instead approximate posterior with a parameterized distribution, Q, with an additional desideratum (mutual information constraint):

$$\begin{aligned}&\min_{\theta_f} \ \mathbb{E}_{\tilde{x}\sim P_X(\cdot)}[\mathbb{E}_{\tilde{n}\sim Q(\cdot|f(\tilde{x}))}[\log Q(\tilde{n}|f(\tilde{x}))]], \\ &s.t.\ Q(n(X)|f(X)) = P(n(X)|f(X)).\end{aligned} \tag{5}$$

The objective is directly calculated with Q in Eq. (5). Hence, the backbone network f can be trained under the additional desideratum with no change to the basic training procedure.

It is difficult to calculate or optimize while satisfying the constraint in Eq. (5). The intuitive meaning of mutual information constraint is clear: the smaller the KL divergence between P and Q, the greater the closeness between Q and P, indicating that Q gets more information from P as learning gradually continues. Therefore, an approximation of the posterior distribution, the parameterized model Q, can be achieved through KL divergence. Modeled with tractable distribution, the novel regularization loss, $\mathcal{L}_{IT}$, can be written as follows:

$$\begin{aligned}\mathcal{L}_{IT} = &\mathbb{E}_{\tilde{x}\sim P_X(\cdot)}[\mathbb{E}_{\tilde{n}\sim Q(\cdot|f(\tilde{x}))}[\log Q(\tilde{n}|f(\tilde{x}))]] \\ &+ \mu D_{KL}(P(n(X)|f(X))||Q(n(X)|f(X))),\end{aligned} \tag{6}$$

where D_{KL} denotes the KL divergence and μ is the balancing parameter for the two terms. We can instead approximate the auxiliary distribution, Q, with the non-LP RPN network h, thus the KL divergence in Eq. (6) is minimized. Approximating $P(n(X)|f(X))$ with the additional network h will minimize D_{KL}, making the problem in Eq. (6) tractable.

By making $D_{KL}(P(n(X)|f(X))||Q(n(X)|f(X)))$ as small as possible, we employ the cross-entropy loss between $n(X)$ and $h(f(X))$ with parameters θ_f, θ_h. Here, loss of the additional network h in operation $h \circ f$ can be obtained as

$$\mathcal{L}_{\mathcal{N}}(\theta_f, \theta_h) = \mathcal{L}_{hcls} - \mathcal{L}_{hloc}, \tag{7}$$

where $\mathcal{L}_{hcls}$ w.r.t. θ_{hcls} and $\mathcal{L}_{hloc}$ w.r.t. θ_{hloc} are classification and localization losses for h RPN sub-networks, respectively. We note that the mutual information term in Eq. (3) is related to classification and not to sophisticated localization. For example, embedded features extracted via f rely heavily on non-LP's classification features, regardless of the results of localization. In an extreme case, even if localization is inaccurate, it is enough to perceive only non-LP information in the image.

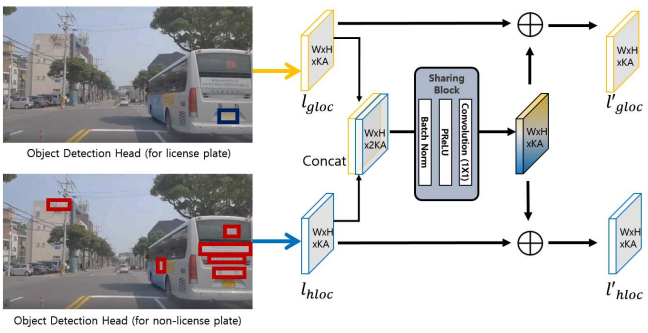


Fig. 3. **Illustration of localization refinement process.** Each localization sub-network in detection head calculates the last feature map l_{gloc} and l_{hloc} respectively considering all proposal boxes. It is utilized into l'_{gloc} and l'_{hloc} using a sharing block on the concatenation of each last feature in addition to the identity feature. Best viewed on the computer, in color and zoomed in.

We can rewrite the formulation of Eq. (6) by relating to Eq. (7) in an adversarial manner. Ideally, the LP-invariant features of f should confuse h which aims at detecting the non-LP. Conversely, the f leverages a model g to detect the only LP by minimizing the detection loss. Namely, we adopt a minimax problem on the θ_f and θ_h, encouraging f to encode only LP-specific visual features into the representations, in which case the classification capability of the non-LP might be harmful. Here, we define the last D_{KL} term of Eq. (6) as the $\mathcal{L}_{IT}$ and can be rewritten as follows:

$$\begin{aligned}\min_{\theta_f}\max_{\theta_h} &\mathbb{E}_{\tilde{x}\sim P_X(\cdot)}[\mathbb{E}_{\tilde{n}\sim Q(\cdot|f(\tilde{x}))}[\log Q(\tilde{n}|f(\tilde{x}))]] \\ &- \mu\mathcal{L}_{\mathcal{N}}(\theta_f, \theta_h).\end{aligned} \tag{8}$$

Specifically, we train the detection network to minimax Eq. (3) by alternating information theory terms into Eq. (8), and primal detection loss can be further expressed as:

$$\begin{aligned}\min_{\theta_f\theta_g}\ \max_{\theta_h}\ &\mathcal{L}_{gloc} + \mathcal{L}_{gcls} \\ &+ \lambda_{obj}\mathbb{E}_{\tilde{x}\sim P_X(\cdot)}[\mathbb{E}_{\tilde{n}\sim Q(\cdot|f(\tilde{x}))}[\log Q(\tilde{n}|f(\tilde{x}))]] \\ &- \mu(\mathcal{L}_{hcls} - \mathcal{L}_{hloc}).\end{aligned} \tag{9}$$

Optimizing this loss function requires adversarial learning strategy [48], [49] of the networks, f, g and h. In addition, we apply gradient reversal layer (GRL) [50] after $f(X)$.

D. Localization Refinement Module

In order to make the regressed bounding box coordinates by the localization sub-network easier to predict, we also introduce the process of localization refinement. To provide the complementary information of the bounding box in the training process, we employ a sharing block $\mathcal{S}(\cdot)$ for refining the localization feature.

We are given a set of feature maps l by the localization sub-networks, where $\{l = [l_{gloc}, l_{hloc}]\}$ contains the last feature maps for g_{loc} and h_{loc} respectively. Then, l is fed into the proposed $\mathcal{S}$ for the localization refinement and output l', where $\{l' = [l'_{gloc}, l'_{hloc}]\}$ the refined feature maps corresponding to g_{loc} and h_{loc} respectively. Figure 3 shows the process of

localization refinement. The architecture for the sharing block $\mathcal{S}$ of the localization information follows three consecutive operations: Batch Normalization (BN) [51], followed by a PReLU [52] activation function and a 1×1 convolution layer. The sharing block $\mathcal{S}$ connects the concatenated feature map between the last feature map of the localization sub-networks, l_{gloc} and l_{hloc}, respectively. This gives rise to the following layer transition: $l_{\mathcal{S}} = \mathcal{S}(l_{gloc}, l_{hloc})$, where $l_{\mathcal{S}}$ denotes the output of the $\mathcal{S}$. Motivated by [47], we add a skip-connection between the output of the sharing block and the last feature map l in each localization sub-network:

$$l' = \mathcal{S}(l_{gloc}, l_{hloc}) + l. \tag{10}$$

Our refinement module plays two roles, where the first is to complement each localization information of sub-network by maximizing the opportunities for useful conjunctions. In fact, optimizing Eq. (9) will make the localization network h_{loc} of the non-LP detector h more stable by $\mathcal{L}_{\mathcal{N}}$. That is, h_{loc} is likely to have the ability to accurately locate not only LP, but also scene-text that looks like LP. Thus, it is likely to complement g_{loc}. The other role is to promote the localization sub-networks to regress precise objects.

E. Training

A pre-trained CNN model [53] is employed as the backbone network. For stable gradient calculation, we optimize the objective function Eq. (9) in an alternative way [48], [54] instead of a straightforward way and the modified optimization objective in terms of $g \circ f$ and $h \circ f$ can be represented as Eq. (11) and Eq. (12) respectively:

$$\min_{\theta_f \theta_g} \mathcal{L}_{gloc} + \mathcal{L}_{gcls}, \tag{11}$$

and

$$\min_{\theta_f \theta_h} (-\lambda_{obj} \mathbb{E}_{\tilde{x} \sim P_X(\cdot)} \mathbb{E}_{\tilde{n} \sim Q(\cdot|f(\tilde{x}))} [\log Q(\tilde{n}|f(\tilde{x}))]) + \mu(\mathcal{L}_{hcls} - \mathcal{L}_{hloc}). \tag{12}$$

At the beginning of training, $g \circ f$ was trained to detect the LP including non-LP information. h from feature extractor with non-LP information also learned to detect non-LP adequately. As the learning progresses, f is led to extract as much LP-specific features excluding non-LP information as possible, and the h increasingly struggles to detect non-LP because f gradually leverages to make h a poor performing network. At the end of learning, f extracts only LP-invariant feature embedding while ignoring non-LP information completely, given enough capacity. Due to the embedded f, g detects only LP and h is guided to the detector with poor performance, as shown in Fig 4. Further analysis on the proposed method is presented in Section. V.C-E.

F. Inference

At the testing phase, the $h(\cdot)$ task is removed. Given a test image X_{test}, the $g \circ f$ output is the detection result via feature extractor f and LP detection network g. Then, the output result is represented as LPR_{result} follows:

$$LPR_{result} = g(f(X_{test})). \tag{13}$$

IV. New Benchmark: LPST-110K

There are many datasets of LP detection [33], [35], [41], [43], [44] which are available mainly for LP detection. However, these datasets do not provide annotation of the scene text (not LP) bounding box.

We collected images of LP and scene-texts to make the new dataset and the benchmark. The dataset is focused on images taken from moving and static cameras as it is meant to be useful for real-world applications. LPST-110K collected images from hundreds of dash and surveillance cameras are being mounted in driving vehicles and building respectively, including locations in East Asia and Europe. We include the scene texts, such as non-LP (e.g. traffic sign, wallpaper text, banner, commercial advertisements, *etc.*), and also includes LP. By doing so, we do not restrict that the instances are taken from the uncontrolled settings (Table I). Each correctly detected scene texts is captured in 5 images, as it is passing by the camera or themselves. The dataset contains 110,000 scene text instances of 9,795 images. The scene texts are divided into two classes: 51,031 LP instances and 58,969 non-LP instances. The properties in the dataset are shown in Table I and samples from the dataset are in Figure 5-7 and 9-10. The data include information about the 2D bounding box for each instance and recognition annotation with letters extracted manually.

Our proposed dataset is very challenging in diverse ways: density, image quality, illumination, angle, distance and complex background, and so on. For example, density (How objects densely indicated in image?, LP/LP + nonLP) is closest to real-world scenarios, that frequently appear on the scenes of all images. We reflect such property to LPST-110K as follows: *AOLP* - 1/1, *SSIG* - 4.34/4.34, *UFPR* - 1/1, *CD-HARD* - 1/1, *CCPD* - 1/1, ***LPST-110K* - 5.21/11.00**. Besides, our dataset is also unique and difficult due to the existence of non-LP, because their presence is the biggest obstacle to LP detection. As we analyze, the non-LP instance will cause more false-positive errors. The resolution of each image is 1280 (Width) $\times$ 720 (Height) $\times$ 3 (Channels). Specifically, this resolution is enough to leverage LP-related tasks. Also, the images in LPST-110K are compressed by h264 codec setting, and unlike most existing LP detection datasets, our tilt degrees, distance, illumination, and blur degrees are diverse and not just frontal or rear. LPST-110K is representative of real-world scenarios where LP detection may be desired.

V. Experiments

A. Implementation Details

All the reported implementations are based on Pytorch as learning framework, and the method was done on the NVIDIA TITAN X GPU and one Intel Core i7-6700K CPU. For stable training, we use a gradient clipping trick and the Adam optimizer [55] with a high momentum. All models are trained for the first 10 epochs with a learning rate of 10^{-4}, next 11-20 epochs with the learning rate of 5×10^{-5}, and then for the remaining epochs at the learning rate of 10^{-5}. For f, we used the ResNet-50 as the backbone, which is pre-trained on ImageNet [53] except for the last fully connected layer. It was then fused with the upsampled result from the deeper

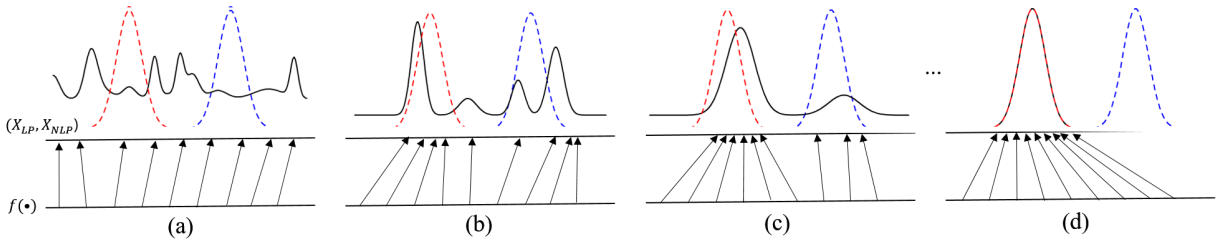


Fig. 4. **The training process of $g \circ f$.** $g \circ f$ (black, solid line) are trained to detect the LPs using $f(\cdot)$ as input so that it can classify between samples from the LP data distribution (red, dotted line) and non-LP data distribution (blue, dotted line). The horizontal line below is the feature extraction from which f is sampled. The upper horizontal line is part of the multi-data distribution of X_{LP}(LP data distribution) and X_{NLP} (non-LP data distribution). The upward arrows indicate how the mapping $(X_{LP}, X_{NLP}) = (g \circ f)$. (a) The initial state before learning randomly is mapped regardless of the distribution of the data. (b) At the beginning of the training, $(g \circ f)$ learns both LP and non-LP information. (c) After several steps of training, $(g \circ f)$ will be guided to intensively learn LP and will gradually ignore non-LP. (d) Lastly, at the end of the training, the LP distribution will reach a point at which sampled LP data distribution because it is learned to ignore non-LP information.

FPN layer. Finally, we apply a 3×3 on a 256 feature size convolutional layer with the same padding as the feature for object detection. Subsequently, this applies two additional 3×3 on 256 feature size, /2 convolution on the deepest layer of the backbone to detect extremely large objects.

For classification sub-networks (g_{cls} and h_{cls}) and localization sub-networks (g_{loc} and h_{loc}), a fully convolutional network is employed, consisting of four times 3×3 on 256 feature size convolutional layers with the same padding and PReLU [52] activation. Each sub-network is trained with CCE loss [56] for classification and L1 smooth loss [2] for 4-axis box coordinates regression. The experimental results are presented in the following sections.

B. Datasets and Evaluation Metrics

We test our method on five LP detection benchmarks AOLP [43], UFPR [33], PKU [45], CCPD [41] and newly collected dataset, named LPST-110K. The first four benchmarks are collected for addressing license plates, while the last one targets at providing not only LP but also non-LP scene text. In existing datasets, all except LPST-110K are the annotated dataset only for LP. Since non-LP detection network h requires non-LP data, we initially train the proposed model using only LPST-110K except them. To provide more kind comparisons for its performance, we also retrain $g \circ f$ during freezing h using existing datasets.

AOLP [43] can be split into three categories: AC, LE and RP. Testing images of each subset consist of 581, 757, and 611 images.

UFPR [33] images are partitioned into train, validate, and test splits. Training consists of 50% of the images (1,800 images); 20% of the images (900 images), are used for validation. The rest, 1,800 images is used for testing.

PKU [45] images are captured in daylight (G1), daylight with sun-glare (G2), nighttime (G3), nighttime with reflective glare (G4). It provides 3,977 images and 4,389 LP instances.

CCPD [41] consists of 150K images for testing. Most images in this dataset are extremely distorted.

LPST-110K contains 9,795 images and their associated 110,000 scene text bounding boxes, which are divided into 5,795/4,000 images for training and testing, respectively. In addition, LP and non-LP instances consist of 29,891/29,078 and 21,065/29,966 bounding boxes (training/testing), respectively.

Evaluation Metrics As for our proposed model, precision, recall, F-measure, AP are utilized as evaluation protocols. For AOLP, UFPR, CCPD benchmarks, we employ precision and recall metrics that have been widely used in LP detection evaluation. Define precision as:

$$\text{Precision} = \frac{T_p}{T_p + F_p}, \tag{14}$$

where T_p and F_p are the correctly estimated bounding box and the incorrectly estimated bounding box. The precision is the ratio of the quantity of the correctly detected bounding boxes among all the acquired bounding box candidates. The more the detection network produces more non-GT bounding boxes as true positives, it will acquire higher precision.

Define recall as:

$$\text{Recall} = \frac{T_p}{T_p + F_n}, \tag{15}$$

where F_n is the quantity of the undetected ground truth. The recall is the ratio of the correctly estimated bounding boxes among all the ground truths. The more the detection network fails to detect the GT bounding box, the lower the recall.

The IoU is defined as follows:

$$\text{IoU} = \frac{\text{area}(R_{\text{det}} \cap R_{\text{gt}})}{\text{area}(R_{\text{det}} \cup R_{\text{gt}})}, \tag{16}$$

where R_{det} and R_{gt} are area of the detected bounding box and the ground truth respectively. The detected bounding box is considered correct when its IoU overlaps the ground truth region by more than 50% (IoU > 0.5).

In addition, we adopt an F-measure that has been used in the PKU benchmark for LP detection evaluation. The F-measure is calculated as follows:

$$\text{F} - \text{measure} = \frac{2 \times (\text{Precision} \times \text{Recall})}{(\text{Precision} + \text{Recall})}. \tag{17}$$

For LPST-110K, we adopt the average precision (AP) at IoU = .50:.05:.95 (standard challenge metric) and AP at IoU = .75, $\text{AP}^{.75}$ (STRICT LP detection metric).

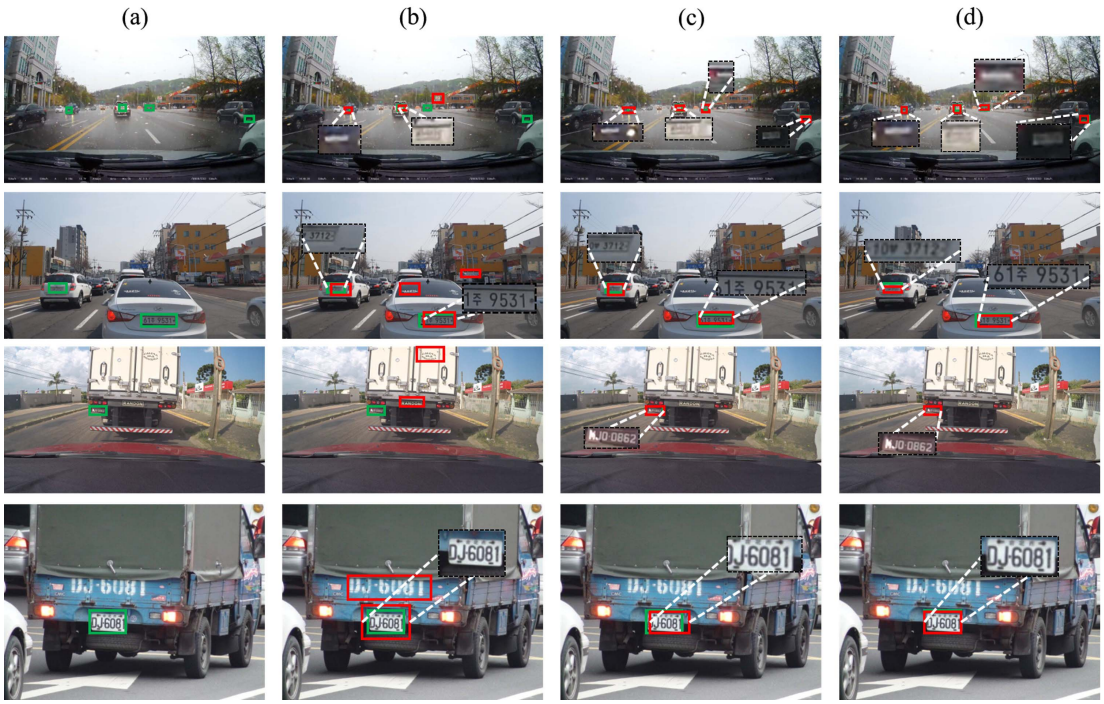


Fig. 5. **Qualitative Results on LPST-110K (first rows), UFPR (second row), and AOLP (last row).** Green bounding boxes are ground truth annotations of LP and red bounding boxes are the detection results. **(a)**: input image; **(b)**: detection result in baseline; **(c)**: adds information loss to (b); **(d)**: adds localization refinement module to (c) - *namely ours*. Best viewed on the computer, in color and zoomed in.

Fig. 6. **Ablation Qualitative Results on LPST-110K.** Best viewed on the computer, in color and zoomed in.

C. Comparisons With State-of-the-Art Methods

For the AOLP, PKU, UFPR, CCPD and LPST-110K, our proposed method can significantly improve the performance of detection, including challenging real-world images as shown in Fig 5 and 7. The results assure that our method consistently enhances the LP detection performance in various datasets. For the AOLP dataset, Table III shows that precision and recall values are nearly as accurate as recent methods. In AOLP, our method generally outperforms the existing state-of-the-art methods. In Table III, [59] has partially better results than our method (e.g. 100 vs 99.71 in the AC subset Precision). However, [59] creates very unrealistic synthetic images that cannot be found in a typical traffic scene to improve this performance, which consists of 450,000 images. In AOLP, using 450,000 datasets for a slight performance improvement requires excessive training time and is inefficient than our method in terms of hardware efficiency. More importantly, our approach leads to better performance in precision, which implies that our method decreases the false positive error regardless of non-LP. This indicates that our method is most suitable as a backbone for our approach both in terms of performance and hardware.

Table IV summarizes the performance of the detection improvement of our approach over the baseline on the three datasets. Specifically, our method obtains the highest performance **(99.17%)** and **(96.1%)** in UFPR and CCPD, and

Fig. 7. **Qualitative detection Results on LPST-110K (Challenging samples).** The first two rows are images taken from the Driving View and include license plates with unusual positions or very tiny size. The samples in the last row are images taken from the drone view, and the angle size variations are very frequent. Red bounding boxes are LP detection results and yellow bounding box non-LP results. Best viewed on the computer, in color and zoomed in.

TABLE III

COMPARISON OF DETECTION RESULTS BY DIFFERENT METHODS ON THE AOLP DATASET

Method/Subset	AC		LE		RP	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Hsu *et al.* [43]	91	96	91	95	91	94
Li *et al.* [39]	98.53	98.38	97.75	97.62	95.28	95.58
Selmi *et al.* [57]	92.6	96.8	93.5	93.3	92.9	96.2
Rafique *et al.* [20]	-	98.09	-	93.92	-	89.03
Xie *et al.* [19]	99.51	99.51	99.43	**99.43**	99.46	**99.46**
Li *et al.* [58]	-	99.12	-	99.08	-	98.2
Björklund *et al.* [59]	**100**	99.3	**99.8**	99.0	**99.8**	99.0
Selmi *et al.* [60]	99.3	**99.4**	99.2	99.2	98.9	98.8
Ours	**99.71**	**99.8**	**99.9**	**99.47**	**99.83**	**99.46**

TABLE IV

COMPARISON OF DETECTION RESULTS BY DIFFERENT METHODS ON THE PKU, UFPR AND CCPD DATASETS

Method/Dataset	PKU (Accuracy)						UFPR	CCPD
	G1	G2	G3	G4	G5	Average	Recall	Precision
Faster RCNN [2]	-	-	-	-	-	-	-	92.9
SSD300 [5]	-	-	-	-	-	-	-	94.4
YOLO9000 [61]	-	-	-	-	-	-	-	93.1
RetinaNet (Baseline) [4]	96.67	97.29	96.77	96.68	95.34	96.34	97.22	94.1
Zheng *et al.* [62]	94.93	95.71	91.91	69.58	67.61	79.47	-	-
Wang *et al.* [63]	-	-	-	-	-	-	-	47.2
Zhao *et al.* [64]	95.18	95.71	95.13	69.93	68.1	80.29	-	-
Zhou *et al.* [65]	95.43	97.85	94.21	81.23	82.37	90.22	-	-
Li *et al.* [66]	98.89	98.42	95.83	81.17	83.31	91.52	-	-
Yuan *et al.* [45]	98.76	98.42	97.72	96.23	97.32	97.69	-	-
Li *et al.* [58]	**99.88**	**99.86**	99.60	**100**	99.31	99.73	-	94.2
Xu *et al.* [41]	-	-	-	-	-	-	-	94.5
Laroca *et al.* [33]	-	-	-	-	-	-	98.33	-
Björklund *et al.* [59]	98.77	99.00	98.92	97.74		98.61	-	-
Laroca *et al.* [38]	-	-	-	-	-	-	98.67	-
Selmi *et al.* [60]	99.5	99.4	99.4	99.6	99.1	99.4	-	-
Ours	**99.88**	**99.86**	**99.87**	99.65	**99.58**	**99.74**	**99.17**	**96.1**

outperforms other state-of-the-art methods by more than 0.5% and 1.6%. Partially, the performance in PKU is lower than other method [58] (e.g. 100 vs 99.65 in G4 subset) However, in all subsets except for the G4 subset, our method outperforms the others, even on the overall average. In addition, in the more unrestrained and challenging UFPR and CCPD, the performance outperforms any other methods. Please note that UFPR and CCPD are much more challenging than PKU. UFPR and CCPD are more diverse and complex in terms of both geometric and semantic views. It is worth addressing that

TABLE V
COMPARISON OF DETECTION RESULTS BY DIFFERENT BACKBONE NETWORKS ON AOLP, PKU, UFPR, CCPD AND LPST-110K DATASETS

Backbone/Dataset	AOLP						PKU	UFPR	CCPD	LPST-110K	
	AC		LE		RP		Average	Recall	Precision	AP	AP.75
	Precision	Recall	Precision	Recall	Precision	Recall	Accuracy				
VGG16bn [67]	98.81	98.97	98.46	98.02	98.43	98.69	99.70	96.56	90.80	.850	.759
ResNet-18	92.60	95.18	98.34	97.62	99.20	98.20	99.45	95.56	89.50	.841	.718
ResNet-34	**99.71**	99.66	99.65	98.68	99.37	99.02	99.70	98.61	94.13	.880	.857
ResNet-50	**99.71**	**99.80**	**99.90**	**99.47**	**99.83**	**99.46**	99.74	**99.17**	**96.10**	**.911**	**.915**
ResNet-101	**99.71**	99.66	99.45	99.32	99.35	**99.79**	99.79	99.06	95.00	**.911**	.904
ResNet-152	**99.71**	**99.80**	99.40	99.32	98.40	**99.79**	99.79	**99.17**	96.05	.872	.911
DenseNet121 [30]	92.65	96.73	97.35	98.02	94.01	96.56	99.32	90.78	85.19	.869	.844
ResNext50 ResNext50 [68]	**99.71**	**99.80**	99.12	99.32	99.32	**99.79**	**99.82**	99.11	95.29	.907	.887
MobileNet-v2 [69]	99.30	**99.80**	98.33	98.28	97.96	99.18	99.48	95.78	91.47	.849	.701

the new method can benefit from the proposed information loss because it prevents non-LP detection even the wild scenes.

Table V reports the results for the newly collected LPST-110K. Still, we can see the same pattern that our method non-trivially increases detection accuracy in both experiments: 1) targeted only LP and 2) targeted all of scene texts. Our approach robustly improves the performance regardless of the presence of non-LP as shown in Figure 5-7.

D. Ablation Study

We perform an ablation study about the effect of the proposed information-theoretical loss and localization refinement module. In the baseline, the results of detection often find the non-LP objects. On the other hand, our approach can improve detection performance, because it provides LP-invariant features around unconstrained scenes. Table V shows how much detection accuracy is improved by the proposed method with ablation manner. When employing information-theoretical loss and localization refinement module (LRM) to the baseline, the LP detection performance is further improved by 0.42% and 0.48%. Especially, GRL [50] is used in both LP and non-LP modules before the feature extraction network f. Although the GRL was originally proposed to solve domain discrimination problem, we obtained the performance improvements. Figure 5 and 6 shows the qualitative results. Consequently, all the components improves LP detection performance noticeably, and clearly ignores non-LP information.

To further investigate the effect of the proposed model, we apply the non-LP detection condition to identify the information-theoretical loss from affecting the avoidance of non-LP. The results are shown in the last column (non-LP) of Table V and Figure 5-6. Surprisingly, the precision and recall decrease by 17.1% and 16.8% compared to the baseline. In addition, Figure 8 shows a PR curves on LTSP-110K with $AP^{.75}$, which demonstrate our method proves that each of our components is more effective than the baseline. These results assure that both modules are profitable.

E. Model Analysis

We discuss some model analysis, including "LP recognition results," "Error study," and "the impact of additional network," are discussed in the following:

1) LP Recognition Results: The LP detection and recognition (LPDR) task aims at assessing the overall, end-to-end,

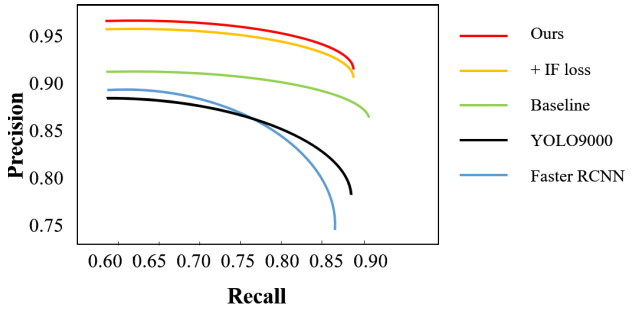


Fig. 8. **PR Curves on LTSP-110K with IOU = 0.75 (*higher curve is better.*)**

LPR system performance. For this task, we define a true positive LP detection and recognition as 1) the LP has been precisely localized within the image with IoU $>$ 0.5 and 2) all the characters in the LP have been precisely recognized. The LPDR performance is also measured in terms of accuracy, as defined in the LP detection task.

For character recognition (CR), we utilize a CNN-LSTM encoder and decoder. In the encoder, the input is an output from the proposed detector. In the same vein, the area of the LP is mostly very small relative to the input image. Therefore, only seven lower convolutional layers of the encoder are used to extract features with two 2×2 max-pooling operations. The encoder network is followed by Bi-directional LSTM [70] each of which uses 256 hidden units that explicitly control data flow. For the decoder, we employ the attentional mechanism with GRU [71] and LSTMs. In the inference phase, the decoder predicts an individual text class y_k at step k until the last step of scene text, where k is the number of predicted characters. Additionally, we show the LPDR results of the images acquired on the LPST-110K as shown in Fig 10.

The AOLP [43] dataset is challenging because the LP's angle contains oblique samples in terms of distortion. On the other hand, in terms of resolution, all images are relatively easy to recognize because they consist of high resolution samples rather than other datasets. Throughout the experiments, we compared our method with other state-of-the-art LPR methods. Overall, our method obtains the highest performance **(97.36%/99.09%/98.63%)**, and outperforms others in LE and RP subsets.

PKU and UFPR datasets samples are far from the camera, causing an issue in terms of resolution. However, they are

TABLE VI

LPDR PERFORMANCE. FULL LPR PERFORMANCE (PERCENTAGE) COMPARISON OF OUR METHOD WITH THE EXISTING METHODS. IN [43] THE AUTHORS PROVIDED AN ESTIMATIVE, AND NOT THE REAL EVALUATION. BEST RESULTS IN EACH CATEGORY ARE IN **BOLD**

Method/Type	AOLP			PKU					UFPR
	AC	LE	RP	G1	G2	G3	G4	G5	
Christos *et al.* [72]	92	86	91	-	-	-	-	-	-
Jiao *et al.* [73]	90	86	90	-	-	-	-	-	-
Hsu *et al.* [43]	86.6	83	85.7	-	-	-	-	-	-
Li *et al.* [39]	94.85	94.19	88.38	-	-	-	-	-	-
Selmi *et al.* [57]	96.2	95.4	95.1	-	-	-	-	-	-
Li *et al.* [58]	95.59	96.43	83.8	-	-	-	-	-	-
Silva *et al.* [35]	-	-	98.36	-	-	-	-	-	-
Björklund *et al.* [59]	94.6	97.8	96.9	-	-	-	-	-	-
Laroca *et al.* [33]	-	-	-	-	-	-	-	-	64.89
Laroca *et al.* [38]	-	-	-	-	-	-	-	-	**82.5**
Selmi *et al.* [60]	**97.8**	97.4	96.3	**98.4**	98.0	97.8	**97.3**	96.9	-
Baseline [4] + CR	97.06	97.11	94.72	92.96	93.71	94.08	94.41	93.53	80.83
Ours + CR	97.36	**99.21**	**98.85**	98.02	**98.71**	**97.85**	96.15	**97.08**	81.67

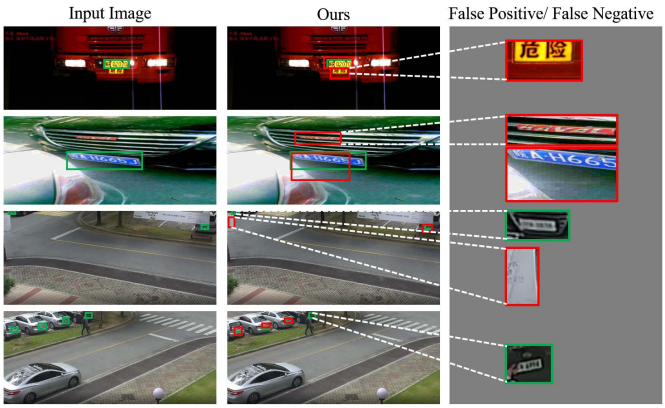


Fig. 9. **Error study on PKU (first row), CCPD (second row), and LPST-110K (3rd-4th rows) dataset.** In the first column, Green bounding boxes are ground truth annotations of LP. In the second column, Red bounding boxes are our detection results. In the last column, the red bounding boxes are false-positive errors and the green bounding boxes are false-negative errors.

almost invariant in terms of distortion because the captured environments are hardly affected by the tilted LP angle or lighting. Under such conditions, the proposed method achieves a competitive performance over most state-of-the-art LPR methods, as shown in Table VI. Specifically, we note the role of localization refinement module, where tiny-LPs often appear in these dataset, and are likely to be unclassified as non-plates because they contain minimal pixel information. Nevertheless, our method produces high-performing localization that can be further adapted from LP, thereby reducing the false-positive and false-negative error. In Table VI, last two rows (baseline and ours) show that results of our method.

2) Error Study: We tested our approach on LPST-110K and four existing benchmarks for LP detection, and show how it to surpasses existing detection methods achieving remarkable performance. However, even the best results on LPST-110K are far from being saturated, suggesting that these unconstrained scenes remain a challenging frontier for future work. Figure 9 shows some cases of failure, including some false recognition results. These results identify that more progress is needed to further improve detection performance. From Figure 9, it can be observed that the overall imaging conditions are low-quality images collected in unconstrained environments. For example, the image in the first row contains uneven illumination from the night and image in the second row is taken at very tilted angle. Specifically, the cases of the LP images in the 3rd to 4th rows are captured at very low-resolutions.

The probable causes of failure include low-quality images and severe interferences. In the first row, a false-positive error occurred, and they have a background and form very similar to LP. Then, since the LP in the second row is very tilted and low quality, not only did it fail to detect correctly, but it also caused another false detection by the logo. Finally, the last two rows show false detection due to banners and occlusion. Considering the failure cases of errors, most errors can be solved by prior knowledge related to text recognition information, and if not, our proposed method is almost close to the human-level.

3) Impact of Additional Network: In this section, we further perform experiments to analyze the performance of our proposed method. We compare the structure of our additional network h with other types of networks to demonstrate the efficiency of a dual network with different purposes. The objective of the LP detector is to detect as many LPs as accurately as possible. Our ultimate goal is to provide the possibility to be able to recognize even the hard positive LPs contained in the unconstrained image. In Table VII, the performance of detection is shown to depend on how the structure is designed. We can see that additional network h with different objectives show better performance among them. The existing method [4] that focuses too much on LPs tends to ignore the characteristics of hard-positive LPs, and does not even provide a chance for recognition (see the Baseline). Most importantly, when a two-class object detector simultaneously detects both LP and non-LP, we can identify that the results exhibit fairly high performance. This implies that the two-class detector can detect LP quite accurately. Although it may work well for us to find the right candidate for the target we want, it still causes too many errors and only shows the same or slightly better performance than our method (24.5%/21.1% and 20.2%/15.3% in IoU = .5 and 22.1%/21.1% and 9.3%/7.7% in IoU = .75). This confirms that

TABLE VII

ADDITIONAL ANALYSIS ON LPST-110K DATASET. BASELINE IS RETINANET [4] WITH RESNET-50 [47]. THE HIGHER LP-RELATED RESULTS, THE BETTER. AND THE LOWER NON-LP-RELATED RESULTS, THE BETTER. BEST RESULTS IN EACH CATEGORY ARE IN **BOLD**

Method Type	LP		non-LP (IoU=.5)		non-LP (IoU=.75)	
	AP	AP75	Precision	Recall	Precision	Recall
Baseline	.863	.769	37.3	32.1	23.1	23.1
Baseline (Two-Class Detector)	.899	.784	24.5	21.1	22.1	21.1
Ours	**.911**	**.915**	**20.2**	**15.3**	**9.3**	**7.7**

Fig. 10. **Qualitative LPDR results of our proposed method.** Green bounding boxes are ground truth annotations of LP and red bounding boxes are the results from our method.

the proposed method can effectively perform discriminative feature learning and filter out unnecessary candidates.

F. Speed

The training speed is about 7.9 iterations/s, taking less than 2 days to reach convergence. In terms of inference, compared to other methods, the proposed model shows a good accuracy-speed trade-off. It is designed for highly accurate LP detection, running at 14 FPS for the input scale 1280 $\times$ 720. Though being a little slower than the fastest method [41], it overcomes [41] accuracy by a large margin. Besides, the speed of ours could be boosted with greater batch size.

VI. CONCLUSION

In a controlled environment, the performance of modern LP detectors is amazing, but still limited. This study focuses on unconstrained real-world scenes, including scene text samples, and provide LPST-110K, a new benchmark for such real-world images, for training and testing with detection annotations. In many emerging state-of-the-art detectors, our experiments on this benchmark show their performance is not guaranteed in a complex environment. To solve this problem, the LPST-110K is used to provide two techniques for robust LP detection in these environments. The first is novel information-theoretical learning that takes advantage of three networks for exploiting LP oriented information. The second technique is a localization refinement for generalizing the bounding box regression network to complement ambiguous detection results. Extensive experiments on diverse benchmarks demonstrated the effectiveness of our method when detecting challenging LPs accurately. This study is helpful for recognition compared to other contemporary approaches.

Future work will address a number of challenging cases identified by this work, in particular the wide variation in how well a combination of text detection and text recognition process improves performance of a license plate detection. Further research could investigate how to complementary connect the text recognition result of a single image to license plate detection, and in turn develop a unified license plate detection and recognition framework.

REFERENCES

[1] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580–587.

[2] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

[3] S. Noh, D. Shim, and M. Jeon, "Adaptive sliding-window strategy for vehicle detection in highway environments," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 2, pp. 323–335, Feb. 2016.

[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.

[5] W. Liu *et al.*, "SSD: Single shot MultiBox detector," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 21–37.

[6] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 779–788.

[7] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[8] M. Veres and M. Moussa, "Deep learning for intelligent transportation systems: A survey of emerging trends," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 8, pp. 3152–3168, Aug. 2020.

[9] J. Yu, D. Y. Kim, Y. Yoon, and M. Jeon, "Action matching network: Open-set action recognition using spatio-temporal representation matching," *Vis. Comput.*, vol. 36, no. 7, pp. 1457–1471, Jul. 2020.

[10] J. Yu, Y. Lee, K. C. Yow, M. Jeon, and W. Pedrycz, "Abnormal event detection and localization via adversarial event prediction," *IEEE Trans. Neural Netw. Learn. Syst.*, early access, Feb. 3, 2021, doi: 10.1109/TNNLS.2021.3053563.

[11] Q. Wang, J. Gao, and Y. Yuan, "Embedding structured contour and location prior in siamesed fully convolutional networks for road detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 230–241, Jan. 2018.

[12] X. Yang, X. Li, Y. Ye, R. Y. K. Lau, X. Zhang, and X. Huang, "Road detection and centerline extraction via deep recurrent convolutional neural network U-Net," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 7209–7220, Sep. 2019.

[13] Q. Wang, J. Gao, and Y. Yuan, "A joint convolutional neural networks and context transfer for street scenes labeling," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 5, pp. 1457–1470, May 2018.

[14] W. Liu, X. Ma, Y. Zhou, D. Tao, and J. Cheng, "p-Laplacian regularization for scene recognition," *IEEE Trans. Cybern.*, vol. 49, no. 8, pp. 2927–2940, Aug. 2018.

[15] J. Gao, Y. Yuan, and Q. Wang, "Feature-aware adaptation and density alignment for crowd counting in video surveillance," *IEEE Trans. Cybern.*, vol. 51, no. 10, pp. 4822–4833, Oct. 2021.

[16] Q. Wang, J. Gao, W. Lin, and X. Li, "NWPU-Crowd: A large-scale benchmark for crowd counting and localization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 6, pp. 2141–2149, Jun. 2021.

[17] Y. Lv, Y. Duan, W. Kang, Z. Li, and F.-Y. Wang, "Traffic flow prediction with big data: A deep learning approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 2, pp. 865–873, Sep. 2015.

[18] Y. Li, R. Yu, C. Shahabi, and Y. Liu, "Diffusion convolutional recurrent neural network: Data-driven traffic forecasting," 2017, *arXiv:1707.01926*.

[19] L. Xie, T. Ahmad, L. Jin, Y. Liu, and S. Zhang, "A new CNN-based method for multi-directional car license plate detection," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 507–517, Feb. 2018.

[20] M. A. Rafique, W. Pedrycz, and M. Jeon, "Vehicle license plate detection using region-based convolutional neural networks," *Soft Comput.*, vol. 22, no. 19, pp. 6429–6440, 2018.

[21] Y. Lee, J. Yun, Y. Hong, J. Lee, and M. Jeon, "Accurate license plate recognition and super-resolution using a generative adversarial networks on traffic surveillance video," in *Proc. IEEE Int. Conf. Consum. Electron.-Asia (ICCE-Asia)*, Jun. 2018, pp. 1–4.

[22] Y. Lee, J. Lee, H. Ahn, and M. Jeon, "SNIDER: Single noisy image denoising and rectification for improving license plate recognition," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1–10.

[23] C. Zhang, Q. Wang, and X. Li, "V-LPDR: Towards a unified framework for license plate detection, tracking, and recognition in real-world traffic videos," *Neurocomputing*, vol. 449, pp. 189–206, Aug. 2021.

[24] S.-Z. Wang and H.-J. Lee, "Detection and recognition of license plate characters with different appearances," in *Proc. IEEE Intell. Transp. Syst.*, vol. 2, Oct. 2003, pp. 979–984.

[25] V. Shapiro, G. Gluhchev, and D. Dimov, "Towards a multinational car license plate recognition system," *Mach. Vis. Appl.*, vol. 17, no. 3, pp. 173–183, Aug. 2006.

[26] T. D. Duan, T. H. Du, T. V. Phuoc, and N. V. Hoang, "Building an automatic vehicle license plate recognition system," in *Proc. Int. Conf. Comput. Sci. RIVF*, vol. 1, 2005, pp. 59–63.

[27] B. Hongliang and L. Changping, "A hybrid license plate extraction method based on edge statistics and morphology," in *Proc. 17th Int. Conf. Pattern Recognit. (ICPR)*, 2004, pp. 831–834.

[28] W. Jia, H. Zhang, and X. He, "Region-based license plate detection," *J. Netw. Comput. Appl.*, vol. 30, no. 4, pp. 1324–1333, Nov. 2007.

[29] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[30] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.

[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[32] S. Kim, H. Jeon, and H. Koo, "Deep-learning-based license plate detection method using vehicle region extraction," *Electron. Lett.*, vol. 53, no. 15, pp. 1034–1036, 2017.

[33] R. Laroca *et al.*, "A robust real-time automatic license plate recognition based on the YOLO detector," 2018, *arXiv:1802.09567*.

[34] S. Montazzolli and C. Jung, "Real-time Brazilian license plate detection and recognition using deep convolutional neural networks," in *Proc. 30th SIBGRAPI Conf. Graph., Patterns Images*, Oct. 2017, pp. 55–62.

[35] S. M. Silva and C. Rosito, "License plate detection and recognition in unconstrained scenarios," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 593–609.

[36] C.-H. Lin, Y.-S. Lin, and W.-C. Liu, "An efficient license plate recognition system using convolution neural networks," in *Proc. IEEE Int. Conf. Appl. Syst. Invention (ICASI)*, Apr. 2018, pp. 224–227.

[37] S.-L. Chen, C. Yang, J.-W. Ma, F. Chen, and X.-C. Yin, "Simultaneous end-to-end vehicle and license plate detection with multi-branch attention neural network," *IEEE Trans. Intell. Transp. Syst.*, vol. 21, no. 9, pp. 3686–3695, Sep. 2020.

[38] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, "An efficient and layout-independent automatic license plate recognition system based on the YOLO detector," 2019, *arXiv:1909.01754*.

[39] H. Li and C. Shen, "Reading car license plates using deep convolutional neural networks and LSTMs," 2016, *arXiv:1601.05610*.

[40] T. K. Cheang, Y. S. Chong, and Y. H. Tay, "Segmentation-free vehicle license plate recognition using ConvNet-RNN," 2017, *arXiv:1701.06439*.

[41] Z. Xu *et al.*, "Towards end-to-end license plate detection and recognition: A large dataset and baseline," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2018, pp. 261–277.

[42] W. Wang, J. Yang, M. Chen, and P. Wang, "A light CNN for end-to-end car license plates detection and recognition," *IEEE Access*, vol. 7, pp. 173875–173883, 2019.

[43] G.-S. Hsu, J.-C. Chen, and Y.-Z. Chung, "Application-oriented license plate recognition," *IEEE Trans. Veh. Technol.*, vol. 62, no. 2, pp. 552–561, Feb. 2013.

[44] G. R. Gonçalves, S. P. G. da Silva, D. Menotti, and W. R. Schwartz, "Benchmark for license plate character segmentation," *J. Electron. Imag.*, vol. 25, no. 5, Oct. 2016, Art. no. 053034.

[45] Y. Yuan, W. Zou, Y. Zhao, X. Wang, X. Hu, and N. Komodakis, "A robust and efficient approach to license plate detection," *IEEE Trans. Image Process.*, vol. 26, no. 3, pp. 1102–1114, Mar. 2017.

[46] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.

[47] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[48] I. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.

[49] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "InfoGAN: Interpretable representation learning by information maximizing generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 2172–2180.

[50] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by backpropagation," 2014, *arXiv:1409.7495*.

[51] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.

[52] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.

[53] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[54] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[55] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.

[56] Y. Kim, Y. Lee, and M. Jeon, "Imbalanced image classification with complement cross entropy," *Pattern Recognit. Lett.*, vol. 151, pp. 33–40, Nov. 2021.

[57] Z. Selmi, M. B. Halima, and A. M. Alimi, "Deep learning system for automatic license plate detection and recognition," in *Proc. IEEE 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 1, Nov. 2017, pp. 1132–1138.

[58] H. Li, P. Wang, and C. Shen, "Toward end-to-end car license plate detection and recognition with deep neural networks," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 3, pp. 1126–1136, Mar. 2019.

[59] T. Björklund, A. Fiandrotti, M. Annarumma, G. Francini, and E. Magli, "Robust license plate recognition using neural networks trained on synthetic images," *Pattern Recognit.*, vol. 93, pp. 134–146, Sep. 2019.

[60] Z. Selmi, M. B. Halima, U. Pal, and M. A. Alimi, "DELP-DAR system for license plate detection and recognition," *Pattern Recognit. Lett.*, vol. 129, pp. 213–223, Jan. 2020.

[61] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7263–7271.

[62] D. Zheng, Y. Zhao, and J. Wang, "An efficient method of license plate location," *Pattern Recognit. Lett.*, vol. 26, no. 15, pp. 2431–2438, 2005.

[63] S.-Z. Wang and H.-J. Lee, "A cascade framework for a real-time statistical plate recognition system," *IEEE Trans. Inf. Forensics Security*, vol. 2, no. 2, pp. 267–282, Jun. 2007.

[64] Y. Zhao, Y. Yuan, S. Bai, K. Liu, and W. Fang, "Voting-based license plate location," in *Proc. 14th Int. IEEE Conf. Intell. Transp. Syst. (ITSC)*, Oct. 2011, pp. 314–317.

[65] W. Zhou, H. Li, Y. Lu, and Q. Tian, "Principal visual word discovery for automatic license plate detection," *IEEE Trans. Image Process.*, vol. 21, no. 9, pp. 4269–4279, Sep. 2012.

[66] B. Li, B. Tian, Y. Li, and D. Wen, "Component-based license plate detection using conditional random field model," *IEEE Trans. Intell. Transp. Syst.*, vol. 14, no. 4, pp. 1690–1699, Dec. 2013.

[67] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

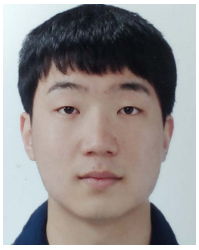
[68] S. Xie, R. Girshick, P. Dollar, Z. Tu, and K. He, "Aggregated residual transformations for deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1492–1500.

[69] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[70] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.
[71] T. Luong, H. Pham, and C. D. Manning, "Effective approaches to attention-based neural machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process.*, 2015, pp. 1–11.
[72] C. N. E. Anagnostopoulos, I. E. Anagnostopoulos, V. Loumos, and E. Kayafas, "A license plate-recognition algorithm for intelligent transportation system applications," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 3, pp. 377–392, Sep. 2006.
[73] J. Jiao, Q. Ye, and Q. Huang, "A configurable method for multi-style license plate recognition," *Pattern Recognit.*, vol. 42, no. 3, pp. 358–369, Mar. 2009.

Younkwan Lee (Student Member, IEEE) received the B.S. degree in computer science from Korea Aerospace University, Gyeonggi, South Korea, in 2016. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea. His current research interests include computer vision, machine learning, and deep learning.

Jihyo Jeon (Student Member, IEEE) received the B.S. degree from the School of Mechanical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, in 2018. She is currently pursuing the M.S. degree with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology. Her current research interests include computer vision, self-driving, and deep learning.



Yeongmin Ko (Student Member, IEEE) received the B.S. degree from the School of Electrical Engineering, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea, in 2017. He is currently pursuing the Ph.D. degree with the School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology. His current research interests include computer vision, self-driving, and deep learning.

Moongu Jeon (Senior Member, IEEE) received the B.S. degree in architectural engineering from Korea University, Seoul, South Korea, in 1988, and the M.S. and Ph.D. degrees in computer science and scientific computation from the University of Minnesota, Minneapolis, MN, USA, in 1999 and 2001, respectively. As the master's degree Researcher, he was involved in optimal control problems with the University of California at Santa Barbara, Santa Barbara, CA, USA, from 2001 to 2003, and then moved to the National Research Council of Canada, where he was involved in the sparse representation of high-dimensional data and the image processing, in July 2005. In 2005, he joined the Gwangju Institute of Science and Technology, Gwangju, South Korea, where he is currently a Full Professor with the School of Electrical Engineering and Computer Science. His current research interests include machine learning, computer vision, and artificial intelligence.

Witold Pedrycz (Life Fellow, IEEE) received the M.Sc., Ph.D., and D.Sc. degrees from the Silesian University of Technology, Gliwice, Poland.

He is currently a Professor and the Canada Research Chair of computational intelligence with the Department of Electrical and Computer Engineering, University of Alberta, Edmonton, AB, Canada. He is also with the Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland. He is a Foreign Member of the Polish Academy of Sciences. He has authored 17 research monographs and edited volumes covering various aspects of computational intelligence, data mining, and software engineering. His current research interests include computational intelligence, fuzzy modeling and granular computing, knowledge discovery and data science, fuzzy control, pattern recognition, knowledge-based neural networks, relational computing, and software engineering.

Dr. Pedrycz is a fellow of the Royal Society of Canada. He was a recipient of the Prestigious Norbert Wiener Award from the IEEE Systems, Man, and Cybernetics Society in 2007; the IEEE Canada Computer Engineering Medal; the Cajastur Prize for Soft Computing from the European Centre for Soft Computing; the Killam Prize; and the Fuzzy Pioneer Award from the IEEE Computational Intelligence Society. He is vigorously involved in editorial activities. He is an Editor-in-Chief of *Information Sciences*, *WIREs Data Mining and Knowledge Discovery* (Wiley), and *International Journal of Granular Computing* (Springer). He currently serves on the Advisory Board for IEEE TRANSACTIONS ON FUZZY SYSTEMS and is a member of a number of editorial boards of other international journals.