SEMANTIC SUPER-RESOLUTION FOR EXTREMELY LOW-RESOLUTION VEHICLE LICENSE PLATE

Yuexian Zou[1,2]*, Yi Wang*[3]*, Wenjie Guan*[1]*, Wenwu Wang*[4]

[1]ADSPLAB, School of ECE, Peking University, Shenzhen, China
[2]Peng Cheng Laboratory, Shenzhen, China
[3]The Chinese University of Hong Kong
[4]Department of Electrical and Electronic Engineering, University of Surrey, U.K.
{zouyx@pkusz.edu.cn, yiwang@cse.cuhk.edu.hk, guanwenjie@pku.edu.cn, w.wang@surrey.ac.uk}

ABSTRACT

Vehicle license plate (VLP) super-resolution (SR) is of great demand in intelligent traffic systems. Super-Resolution for extremely low-resolution VLP remains challenging and the state-of-the-art SR methods hardly provide satisfying results for low-resolution (LR) VLPs. In this study, from a new perspective, we develop an effective solution to achieve the super-resolution of the extremely LR VLP images, by using the semantic information of the characters. Specifically, we firstly exploit the pervasive sparse prior for the character recognition in LR condition for VLPs. Then the semantic information extracted from the sparse representation-based classification (SRC) results is employed to alleviate the illness of the SR problem. To maximize the benefit brought by the semantic information from SRC, we employ sparse-coding based super-resolution (SCSR) method to upscale VLP images. In the end, an exponential soft labeling method is designed to reduce the possible bias introduced by character classification. Extensive experiments on the self-built Chinese VLP dataset (VLP100) and public UFPR-ALPR dataset validate the feasibility and effectiveness of our proposed VLP-SR system.

Index Terms— Super-resolution, vehicle license plate, character recognition, sparse coding, semantic information

1. INTRODUCTION

Super-resolution for vehicle license plates (VLP-SR) is an active research area since it is helpful to identify the target vehicles from the low-resolution (LR) images or videos. VLP-SR plays a vital role in parking lot management, traffic surveillance, finding the stolen vehicles, and so on.

Image super-resolution (SR) is also a hot topic in computer vision, which aims to enhance the image resolution and recover lost high-frequency information for better visual perception. Essentially, SR is an ill-posed problem since there exist countless high resolution (HR) images corresponding to a LR input image. Numerous solutions have been proposed to address this issue. Based on the number of input LR images, SR can be roughly categorized into two types: multi-frame fusion based SR and single image SR (SISR). They both have been redesigned for tackling the VLP-SR problem [1, 2, 3, 4, 5], and achieved improvements in terms of the objective index such as mean square error (MSE), peak signal noise ratio (PSNR), and structural similarity (SSIM).

However, it is noted that, for the VLP-SR problem, the emphasis on the aforementioned objective index does not necessarily leading to promising performance on the later VLP recognition. For example, existing SR and VLP-SR methods may work well for enhancing the VLPs marked by blue in Fig. 1, but most of them fail in enhancing the VLPs marked by red rectangle in Fig. 1. Actually, the characters of VLPs in red are barely recognizable. From the view of visual perception, the high practical demand lies in how to recognize the VLPs marked by red in Fig. 1. Therefore, for such applications, VLP-SR is heavily associated with the identification of the VLPs, which expect the SR-VLP method to produce reliable high resolution (HR) result from an extremely low resolution condition, where the characters of VLPs are barely recognizable. Generally, previous VLP processing systems tend to treat the recognition and SR of the VLPs as two independent procedures [2, 4]. In this work, we unify these two procedures as one, maximize the utilization of semantic priors produced by recognition, and remedy the final SR effects by balancing the bias of priors and the visual performance.

Fig. 1. Example illustration of vehicle license plates obtained by a video camera in the real scenario. The VLPs marked by red rectangle are in the extremely LR condition which are nearly unrecognizable. The VLPs marked by blue rectangle contain intact semantic information although it is unclear.

To address the VLP-SR in the extreme LR condition, we present a novel VLP-SR pipeline by exploiting sparsity and semantics in the VLPs. This work is motivated by the research on semantic super-resolution [6]. By introducing the semantic information to the SR problem, the performance of the original SR methods can be boosted. In this study, the semantic information refers to image segments with corresponding categorical labels. Compared with the natural images, the semantic information in VLPs is more distinctive, since they are merely the limited combination of the English letters and digits.

Specifically, our proposed VLP-SR system exploits the pervasive sparse prior to conduct the character recognition task in LR condition for the VLPs, then uses the semantic information extracted from the sparse representation based classification (SRC) results to alleviate the illness of the SR problem, leading to the clearer SR results. In order to maximize the benefit brought by the semantic information from SRC, we employ sparse-coding based super-resolution (SCSR) method to upscale VLP images. Moreover, since the character classifiers always have bias, the semantic information provided by classifiers will cause potential faults to VLP-SR. To handle such problem, we also design an exponential soft labeling method to reduce such negative influence on SR performance. Finally, to validate the effectiveness of our method, we test our system along with several generic SR methods and redesigned ones on a self-built dataset VLP100 and a public car license dataset UFPR-ALPR.

The remainder of the paper is organized as follows. Section 2 describes the pipeline and the algorithms of our proposed VLP-SR system, and section 3 gives detailed experiments and the corresponding analysis. Finally, section

This paper was partially supported by the Shenzhen Science & Technology Fundamental Research Program (No: JCYJ20160330095814461) & Shenzhen Key Laboratory for Intelligent Multimedia and Virtual Reality (ZDSYS201703031405467). Special acknowledgements are given to Aoto-PKUSZ Joint Research Center of Artificial Intelligence on Scene Cognition & Technology Innovation for its support.

4 concludes the paper.

2. SEMANTIC SUPER-RESOLUTION FOR VLP

Our proposed VLP-SR method aims to exploit the semantics of the VLP to help improve the SR performance. The main task of the VLP-SR is to enhance the visibility of the VLP for improving the recognition of the VLP. With the aforementioned semantics existing in the VLPs, the illness of the SR problem is reduced in the following two ways:

1) The resolution enhancement is directly applied to the semantic object (in this study, the semantic object refers to the image patch of a VLP character) instead of some generic low-level parts (e.g. texture, edges or some other patterns). Thus, the number of potential HR estimations of the given LR images is limited.

2) The aim of the VLP-SR is to help recognize the VLPs, hence VLP-SR should only focus on the information beneficial to the later categorization task. Meanwhile, for the character recognition, only edges matter while the possible texture could be just viewed as stain and the color contains nothing about character recognition. Under this circumstance, without the burden of recovering the texture and some missing color information, the illness of the VLP-SR problem is further alleviated.

Specifically, we employ a discriminative model to infer the semantics of each VLP character in the LR condition, then incorporate the deduced semantics into the later SR task.

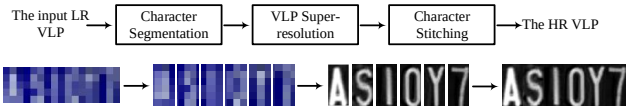


Fig. 2. The pipeline of the proposed VLP-SR method. Above: the processing modules of VLP-SR; below: a real example processed by VLP-SR.

2.1. The pipeline of the proposed method

The pipeline of our VLP-SR system is given in Fig. 2. There are three modules: module 1: LR character segmentation, which divides the VLP image into six sub-images, and each sub-image contains a character; module 2: VLP-SR on the LR character; module 3: HR character stitching, which put the SR character sub-images into a complete HR VLP image. Specifically, module 1 adopts the algorithm proposed in [7], while in module 3, the simple and direct stitching is employed. In the following subsections, the details of our proposed VLP-SR will be discussed.

2.2. The problem formulation of VLP-SR

The scheme of our proposed VLP-SR is summarized as: given a segmented LR character $\boldsymbol{x}$ (in column vector form, $\boldsymbol{x} \in \mathcal{R}^p$), its semantic label $c(\boldsymbol{x})$ is determined by a classifier f. Accordingly, the corresponding HR character $\boldsymbol{y}$ is estimated by

$$\boldsymbol{y} = g(\boldsymbol{x}, c(\boldsymbol{x})) \tag{1}$$

where g is a super-resolution method and $\boldsymbol{y} \in \mathcal{R}^q$. Here $q > p$, and to address the VLP-SR in the extreme LR conditionin, in this study $q \gg p$.

According to the aforementioned problem formulation, the key to VLP-SR is to design a robust classifier f and a proper super-resolution method g. More importantly, f and g should be coordinated properly for boosting the SR performance and reducing the negative effects introduced by the bias of f. In this paper, we employ the sparse representation classification (SRC) [8, 9] to act as f and the sparse coding based SR (SCSR) [10] as g, since SRC works well when given the limited training data [8, 9], and SCSR can exploit the semantic information provided by SRC smoothly as they both utilize the sparsity in image patches.

2.3. The LR-HR semantic dictionary pair constitution

Dictionary learning or constitution is a vital part of the example-based [11, 12, 13] or sparse coding based [8, 9] SR methods. As our g is SCSR, we also need to build our own LR-HR dictionary pair for the VLPs. Let $\boldsymbol{h}_i^j (\boldsymbol{h}_i^j \in \mathcal{R}^q)$ denote the i_{th} HR image of the j_{th} character. Then let $\boldsymbol{D}_h^j$ denote the HR dictionary formed by the HR images of the j_{th} character, where $\boldsymbol{D}_h^j = [\boldsymbol{h}_1^j, \boldsymbol{h}_2^j, ..., \boldsymbol{h}_T^j]$ and the subscript h stands for HR. Correspondingly, the LR dictionary of the j_{th} character consists of the counterpart LR images of the j_{th} character, where $\boldsymbol{D}_l^j = [\boldsymbol{l}_1^j, \boldsymbol{l}_2^j, ..., \boldsymbol{l}_T^j]$ and the subscript l stands for LR. Besides, $\boldsymbol{l}_i^j$ is the down-sampled version of $\boldsymbol{h}_i^j$, which is expressed as: $\boldsymbol{l}_i^j = \boldsymbol{S}\boldsymbol{h}_i^j$, where $\boldsymbol{l}_i^j \in \mathcal{R}^p$ and $\boldsymbol{S} \in \mathcal{R}^{p \times q}$ is the down-sampling operator.

2.4. The VLP-SCSSR algorithm

After getting the character sequence $\boldsymbol{x}^i$ $(i = 1, 2, ..., N)$ from the LR VLPs, we need to acquire the semantic label of each character first. Specifically, with the trained dictionaries $\boldsymbol{D}_l^j$ $(j = 1, 2, ..., K)$, the classification scheme is given as:

Solving the sparse solution of $\boldsymbol{x}^i$ over $\boldsymbol{D}_l^j$:

$$\boldsymbol{w}_j^* = \arg\min_{\boldsymbol{w}_j} \|\boldsymbol{x}^i - \boldsymbol{D}_l^j \boldsymbol{w}_j\|_2^2 \quad s.t. \quad \|\boldsymbol{w}_j\|_1 < t \tag{2}$$

where t is the preset parameter indicating the sparsity of $\boldsymbol{w}_j$.

With $\boldsymbol{w}_j$, the residual of $\boldsymbol{x}^i$ over each dictionary is easily computed as:

$$r_j = \|\boldsymbol{x}^i - \boldsymbol{D}_l^j \boldsymbol{w}_j^*\|_2 \tag{3}$$

Then, the semantic label of each character $\boldsymbol{x}^i$ can be determined based on r_j $(j = 1, 2, ..., K)$. Actually, there are two approaches to design such semantic label:

1) Let $c(\boldsymbol{x}^i)$ be the hard label, meaning $c(\boldsymbol{x}^i)$ is a discrete number, or a vector where 1 represents the target category and 0 for the rest; 2) Let $c(\boldsymbol{x}^i)$ be the soft label, meaning $c(\boldsymbol{x}^i)$ is a vector where each element stands for the confidence to which $\boldsymbol{x}^i$ belongs.

For the first approach, $c(\boldsymbol{x}^i)$ is computed by finding the minimal residual as:

$$c(\boldsymbol{x}^i) = \arg\min_{1 \le j \le K} r_j \tag{4}$$

While for the second approach, the $c(\boldsymbol{x}^i)$ is computed by our definition as:

$$c(\boldsymbol{x}^i) = \frac{1}{Z}[e^{-\alpha r_1}, e^{-\alpha r_2}, ..., e^{-\alpha r_K}] \tag{5}$$

where α is a positive constant and $Z = \sum_{i=1}^{K} e^{-\alpha r_i}$ is the normalization factor. Note that the smaller the residual, the higher the possibility that $\boldsymbol{x}^i$ belongs to the j_{th} character. Actually, we argue that the soft label will reduce the potential faults introduced by the bias of the SRC, since the soft label acts as the weighted sum or the ensemble method, which preserves the possibility of the semantics of $\boldsymbol{x}^i$, and helps increase the representation ability of the SCSR. In the following, we will demonstrate how the soft label works. Assume $\exists r_{t_1} = r_{t_2} = ... = r_{t_M} = \min_{1 \le j \le K} r_j$ and $\min(\{r_j\}_{1 \le j \le K} - \{r_{t_i}\}_{1 \le i \le M}) > r_{t_1}$, then it can be demonstrated that

$$\lim_{\alpha \to +\infty} \frac{1}{Z} e^{-\alpha r_{t_i}} = \frac{1}{M} \quad where\ 1 \le i \le M \tag{6}$$

$$\lim_{\alpha \to +\infty} \frac{1}{Z} e^{-\alpha r_j} = 0 \quad where\ j \ne t_i\ and\ 1 \le j \le K \tag{7}$$

where $M \le K$ (M, K and t_M are all constant). When $M = 1$, it is clear that the soft label reduces to the hard label as only one dimension of $c(\boldsymbol{x}^i)$ is set to 1 and the remaining are 0.

In SCSR, the sparse representation of the LR patch is considered to be consistent with that of the corresponding HR patch over the joint-trained LR-HR dictionary pair, indicating that if $\boldsymbol{x} = \boldsymbol{S}\boldsymbol{y}$ then

$$\begin{cases} \boldsymbol{w}_l^* = \arg\min_{\boldsymbol{w}_l} \|\boldsymbol{x} - \boldsymbol{D}_l \boldsymbol{w}_l\|_2^2 \quad s.t. \quad \|\boldsymbol{w}_l\|_1 < t \\ \boldsymbol{w}_h^* = \arg\min_{\boldsymbol{w}_h} \|\boldsymbol{y} - \boldsymbol{D}_h \boldsymbol{w}_h\|_2^2 \quad s.t. \quad \|\boldsymbol{w}_h\|_1 < t \\ \boldsymbol{w}_l^* \approxeq \boldsymbol{w}_h^* \end{cases} \tag{8}$$

With the formulation of the SCSR, it is clear that the HR estimation $\hat{\boldsymbol{y}}^i$ of $\boldsymbol{x}^i$ is computed as:

$$\hat{\boldsymbol{y}}^i = g(\boldsymbol{x}^i, c(\boldsymbol{x}^i)) \quad with \quad \{\boldsymbol{D}_h^j\}_{j=1,2,...,K} \tag{9}$$

where g is the SCSR in our context.

According to different approaches for determining $c(\boldsymbol{x}^i)$, $\hat{\boldsymbol{y}}^i$ can be computed in the following two different manners:

(a) (b)

Fig. 3. Examples of VLP100 (above) and UFPR-ALPR (below) dataset. (a) The VLP images; (b) the segmented character images.

1) if $c(\boldsymbol{x}^i)$ is hard label, then:

$$\hat{\boldsymbol{y}}^i = \boldsymbol{D}_h^{c(\boldsymbol{x}^i)} \boldsymbol{w}^*_{c(\boldsymbol{x}^i)} \tag{10}$$

2) if $c(\boldsymbol{x}^i)$ is soft label, then:

$$\hat{\boldsymbol{y}}^i = \sum_{j=1}^{K} c(\boldsymbol{x}^i)_j \boldsymbol{D}_h^j \boldsymbol{w}_j^* \tag{11}$$

where $c(\boldsymbol{x}^i)_j = \frac{1}{Z} e^{-\alpha r_j}$.

3. EXPERIMENTS AND RESULTS

3.1. Datasets

We conduct experiments on two VLP datasets, including a self-built Chinese vehicle license plate dataset (VLP100) and the UFPR-ALPR dataset [14], examples are given in Fig. 3. The VLP100 is composed of 100 Chinese VLP images took from the real-world scenarios and with the annotations of their identifications (e.g. "A-635G7", where Chinese characters are omitted). Besides, the UFPR-ALPR dataset is a public large-scale dataset of Brazilian VLPs, containing 4500 images from different types of vehicles (e.g. cars, motorcycles, buses, trucks, among others) with complex backgrounds and lighting conditions.

For these two VLP datasets, we firstly extract two corresponding segmented character datasets from them based on **Module 1** in Section 2.1. The segmented character dataset of VLP100 contains 34 different characters, including 24 English letters ('O' and 'I' are omitted since they are missing from Chinese VLPs) and 10 digits. Each type of the character has 280 RGB images with 48×24 resolution. Similarly, the segmented character dataset of UFPR-ALPR consists of 26 English letters and 10 digits. However, since the images from UFPR-ALPR are acquired with different portable devices (e.g. GoPro Hero4 Silver and iPhone 7 Plus) and affected by vehicle motion, their quality is worse than that in VLP100. We randomly select 300 RGB images from UFPR-ALPR and resized them to 11×7 resolutions for each type of the character.

Then, we preprocess these two VLPs datasets for uniform input. For VLP100, we have 100 cropped and aligned RGB Chinese VLP images with size of around 48×144. They are all taken from the entrance of a parking lot, which are degraded by the disturbance brought by the illumination and environmental noise (see Fig.3). For UFPR-ALPR, 150 VLPs are randomly chosen, cropped, aligned, and resized to 11×49.

In our experiments, since color exerts no influence on the character recognition, we convert all colorful VLPs and character images into gray ones, and they are normalized subsequently.

The semantic dictionary pair is formed by the segmented character data using the approach described in Section 2.3. Since we intend to enhance the visibility of the VLP in the extremely low resolution, the down-sampling factor is set to 8 for VLP100 and 2 for UFPR-ALPR, meaning we prepare to upscale the VLP character images from the resolution of 6×3.

3.2. Performance Evaluation

In this subsection, some benchmark classification algorithms on the VLP character images are evaluated in the LR condition (6×3). Then, VLP-SR experiments have been conducted to show the effectiveness of our proposed VLP-SR method.

Table 1. The classification results on the LR characters of VLP100 and UFPR-ALPR

	KNN	CRC	SRC	MLP	SVM
Err of VLP100	4.41%	13.92%	**2.16%**	4.34%	3.82%
Err of UFPR-ALPR	2.13%	20.18%	**0.08%**	8.14%	12.78%

3.2.1. The character recognition in the extremely low resolution

The experimental results support of using SRC as f are given. We evaluate the performance of different types of classification algorithms in LR VLP character recognition by using the aforementioned two segmented character datasets. Two types of classification methods are used, including: 1) non-parametric methods: K-nearest-neighbors (KNN) [15], sparse representation based classification (SRC) [8, 9], and collaborative representation based classification (CRC) [16]; 2) parametric methods: multiple layered perceptron (MLP) [15] and support vector machine (SVM) [15]. All the used character images are down-sampled to the size of 6×3 for both VLP100 and the UFPR-ALPR. For each type of character, 90% (250 for VLP100 and 270 for UFPR-ALPR) images are randomly selected for training, and the remaining data (30 for both VLP100 and UFPR-ALPR) are for testing.

The classification results on VLP100 and UFPR-ALPR are shown in Table 1. Note that SRC achieves the highest accuracy among all classification methods on both two datasets. Specifically, it achieves 97.84% on VLP100 and 99.92% on UFPR-ALPR, which outperforms the second place method (SVM for VLP100 and KNN for UFPR-ALPR) by around 2% accuracy. It is worth mentioning that 2% higher accuracy suggests SRC is 9.7% higher accuracy than KNN in the license plate recognition task of 7 characters (SRC: $94.5\% = 99.92\%^7$, KNN: $85.82\% = 96.18\%^7$).

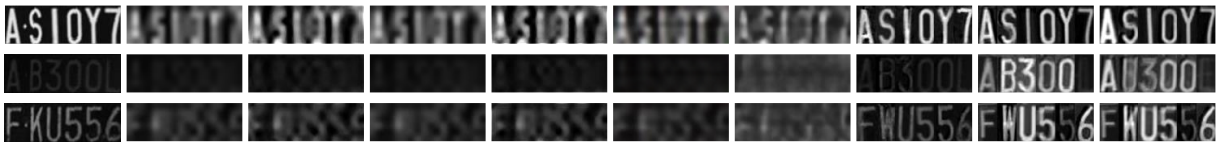
3.2.2. Super-resolution for VLPs in the extremely low resolution

Here, the SR performance of our proposed system and other baseline algorithms are given. The baseline algorithms include: Bicubic interpolation, SCSR [10], Neighborhood embedding (NE) [13], SR convolutional neural network (SRCNN) [17, 18], very deep residual network for SR (VDSR) [19], and semantic super-resolution based on neighborhood embedding (SSR-NE), whose g are NE based SR algorithm while f is still SRC. SCSR, NE, SRCNN, and VDSR are used to show the necessity of incorporating the semantic information into the VLP-SR problem, while SSR-NE is used to show the advantages about maintaining the consistence between f and g. For popular deep learning based SR methods, besides SRCNN and VDSR, we are also aware of various effective SR approaches proposed recently [20, 21, 22, 23, 24]. Almost all of them produce better SR visual effects than SRCNN and VDSR. However, since they are all generic deep learning based algorithms, their VLP SR performance can be expected just like SRCNN and VDSR. Limited training data and image size cannot support the full utilization of the huge model capacity and large receptive field provided by these deep models. The following experiments will verify this argument.

For SRCNN and VDSR, the networks are initialized by the models provided in the source code, and further fine-tuned by 60 training HR VLPs and their corresponding LR ones; for SCSR and NE, such 60 HR-LR VLP pairs are also used for additional training. All the remaining configurations of the compared methods are the same as described in their work. For our proposed VLP-SCSSR, only the semantic dictionary pair mentioned in Section 3.1 is used for training. Additionally, both modes of hard and soft labels are evaluated, and they are named as VLP-SCSSR(h) and VLP-SCSSR(s) for convenience, respectively. In VLP-SCSSR(s), α is set to 80.

SR results on VLP100. The input 100 LR VLP images are generated from the VLP dataset described in Section 3.1. Each one of them is down-sampled at 8 times, which is common in real traffic surveillance applications.

Fig. 4 gives three examples on the SR performance of the methods discussed above. The 1_{st} image shows the SR results of a clear LR VLP image. It is noted that SCSR, NE, A+, SRCNN, and VDSR nearly describe the rough shape of each character, where 'A', 'S', and '1' can be visually recognized while 'Y' and '7' are difficult to infer. In the SSR framework, SSR+NE, VLP-SCSSR(h), and VLP-SCSSR(s) all yield surprisingly visual appealing



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

Fig. 4. Super-resolution results of 3 random images of VLP100. For each result, (a) is the original HR image, (b)-(j) are SR results provided by bicubic interpolation, NE, SCSR, A+, SRCNN, VDSR, SSR+NE, VLP-SCSSR(h), and VLP-SCSSR(s), respectively.

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

Fig. 5. Super-resolution results of 10 random images of UFPR-ALPR. For each result, (a) is the original HR image, (b)-(h) are SR results provided by bicubic interpolation, NE, SCSR, SRCNN, SSR+NE, VLP-SCSSR(h), and VLP-SCSSR(s), respectively.

results. Moreover, SSR+NE has more artifacts than VLP-SCSSR(h) and VLP-SCSSR(s), especially on 'A', '1', and 'Y'. Meanwhile, VLP-SCSSR(h) and VLP-SCSSR(s) have some blurry effects on 'S'.

For the 2_{nd} image, compared with the 3_{th} image, its illumination is poor, in which SCSR, NE, A+, SRCNN, and VDSR give quite blurry results. The compared methods cannot provide any meaningful results. Under such circumstance, SSR+NE, VLP-SCSSR(h), and VLP-SCSSR(s) still offer clearly recognizable results. Besides, since VLP-SCSSR(h) and VLP-SCSSR(s) both share the same sparsity prior and consistence in f and g, in which the illumination problem is treated as noise, their SR results have better contrast compared with SSR+NE. Also, note that VLP-SCSSR(s) (Fig. 4 the 3_{th} image) produces more blurry result compared with VLP-SCSSR(h). Some characters from VLP-SCSSR(s) are obviously the combination of more than two types of characters, such as the second character 'K'. The estimation offered by VLP-SCSSR(s) of 'K' is combined by 'K' and 'W'. From the perspective of human visibility, such blurry combination of character can be annoying. However, on the other side, it alleviates the negative effects introduced by the bias of f. More importantly, SSR+NE and VLP-SCSSR(h) both give clear but semantic wrong estimations as 'FWU556'. In this case, as VLP-SCSSR(s) exploits the confidence of each potential character candidate, its result gives us an opportunity to realize the possibility of the HR ground truth, helping us make a correct decision.

SR results on UFPR-ALPR. Fig. 5 shows the SR performance of the Bicubic interpolation, NE, SCSR, SRCNN, SSR+NE and our proposed method VLP-SCSSR(h) and VLP-SCSSR(s). These 10 example images are randomly selected from the UFPR-ALPR testing set. Each of the LR image has a resolution of 6×3 and its corresponding HR image has a resolution of 11×7.

From the 1_{st}, 2_{nd} and 3_{th} image, we can see that, compared to the results of Bicubic interpolation, SCSR and SRCNN, our method provides clearer SR results. Furthermore, from the 4_{th} to 9_{th} images, we can see that Bicubic interpolation and SRCNN produce blurry results, which are difficult to recognized visually. Additionally, although NE and SCSR generate clear SR results, they give the wrong predict. For example, in the 5_{th} image, NE confuses '4' with 'G' and '6' with '4', while SCSR confuses '4' with 'C'. In the 6_{th} image, NE confuses 'A' with '4', while in the 7_{th} image, NE confuses '8' with '3'. Meanwhile, compared with SCSR and NE, SSR+NE and VLP-SCSSR provide better SR results on similar characters with the semantic information. In the 5_{th} image, the '4' is upscaled to 'C' by SCSR, while VLP-SCSSR provides a correct convincing result as '4'. Moreover, the qualitative comparisons between SSR+NE and VLP-SCSSR validate the necessity of using semantic information and the importance of maintaining the consistence between f and g again.

Considering the input VLPs are in quite low resolution and hard to be recognized, pixel-wise measurements like PSNR or others are not good quantitative evaluation metrics for our desired SR performance. Nevertheless, we give the PSNR values of all used methods and systems for reference in Table 2. Note VLP-SCSSR(h) and VLP-SCSSR(s) only give higher PSNR than SRCNN, and yield much lower PSNR than traditional Bicubic interpolation and SCSR. However, from the perspective of visual qualitative comparison (Fig. 5), VLP-SCSSR produces the clearest VLPs with recognizable characters.

Table 2. Average PSNR on UFRP-ALPR

	Bicubic	SCSR	NE	SRCNN	SSR+NE	VLP-SCSSR(h)	VLP-SCSSR(s)
PSNR (dB)	18.06	20.57	19.02	14.40	19.13	15.0	14.72

4. THE CONCLUDING REMARKS AND FUTURE WORK

We have presented a sparse-coding based super-resolution system with semantic priors for the VLP-SR problem in the extreme LR condition. The unified recognition and SR framework can boost SR effects significantly. Additionally, a soft-labeling method has been developed for producing more convincing SR results. Experiments on our self-built VLP and public UFPR-ALPR datasets show that with our proposed SR system, the original visually unrecognizable VLPs can be clearly displayed.

In the future, we will explore how to seamlessly convert such a system into an end-to-end learning model through neural networks, and explore how to regularize it for improving robustness and dealing with challenging cases.

5. REFERENCES

[1] Shih-Chieh Lin and Chih-Ting Chen, "Reconstructing vehicle license plate image from low resolution images using nonuniform interpolation method," *International Journal of Image Processing*, vol. 1, no. 2, pp. 21–28, 2008.

[2] Louka Dlagnekov, *Video-based car surveillance: License plate, make, and model recognition*, Ph.D. thesis, Citeseer, 2005.

[3] Jie Yuan, Si-dan Du, and Xiang Zhu, "Fast super-resolution for license plate image reconstruction," in *ICPR*. IEEE, 2008, pp. 1–4.

[4] Yushuang Tian, Kim-Hui Yap, and Yu He, "Vehicle license plate super-resolution using soft learning prior," *Multimedia Tools and Applications*, vol. 60, no. 3, pp. 519–535, 2012.

[5] Kaggere V Suresh, G Mahesh Kumar, and AN Rajagopalan, "Super-resolution of license plates in real traffic videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 321–331, 2007.

[6] Radu Timofte, Vincent De Smet, and Luc Van Gool, "Semantic super-resolution: When and where is it useful?," *Computer Vision and Image Understanding*, vol. 142, pp. 1–12, 2016.

[7] J Jagannathan, A Sherajdheen, R Muthu Vijay Deepak, and N Krishnan, "License plate character segmentation using horizontal and vertical projection with dynamic thresholding," in *International Conference on Emerging Trends in Computing, Communication and Nanotechnology*. IEEE, 2013, pp. 700–705.

[8] John Wright, Yi Ma, Julien Mairal, Guillermo Sapiro, Thomas S Huang, and Shuicheng Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, 2010.

[9] John Wright, Allen Y Yang, Arvind Ganesh, S Shankar Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE TPAMI*, vol. 31, no. 2, pp. 210–227, 2009.

[10] Jianchao Yang, John Wright, Thomas S Huang, and Yi M"'a, "Image super-resolution via sparse representation," *IEEE TIP*.

[11] William T Freeman, Thouis R Jones, and Egon C Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[12] Michael Elad and Dmitry Datsenko, "Example-based regularization deployed to super-resolution reconstruction of a single image," *The Computer Journal*, vol. 52, no. 1, pp. 15–30, 2009.

[13] Hong Chang, Dit-Yan Yeung, and Yimin Xiong, "Super-resolution through neighbor embedding," in *CVPR*. IEEE, 2004, pp. I–I.

[14] Rayson Laroca, Evair Severo, Luiz A. Zanlorensi, Luiz S. Oliveira, Gabriel Resende Gon?alves, William Robson Schwartz, and David Menotti, "A robust real-time automatic license plate recognition based on the yolo detector," 2018.

[15] Ethem Alpaydin, *Introduction to Machine Learning*, Pitman, 1988.

[16] Lei Zhang, Meng Yang, Xiangchu Feng, Yi Ma, and David Zhang, "Collaborative representation based classification for face recognition," *arXiv preprint arXiv:1204.2358*, 2012.

[17] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Learning a deep convolutional network for image super-resolution," in *ECCV*. Springer, 2014, pp. 184–199.

[18] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE TPAMI*, vol. 38, no. 2, pp. 295–307, 2016.

[19] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *CVPR*, 2016, pp. 1646–1654.

[20] Christian Ledig, Zehan Wang, Wenzhe Shi, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, and Alykhan Tejani, "Photo-realistic single image super-resolution using a generative adversarial network," in *Computer Vision and Pattern Recognition*, 2017, pp. 105–114.

[21] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 1132–1140.

[22] Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy, "Recovering realistic texture in image super-resolution by deep spatial feature transform," *arXiv preprint arXiv:1804.02815*, 2018.

[23] Ying Tai, Jian Yang, and Xiaoming Liu, "Image super-resolution via deep recursive residual network," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2790–2798.

[24] Wei Sheng Lai, Jia Bin Huang, Narendra Ahuja, and Ming Hsuan Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5835–5843.