Beyond Human-level License Plate Super-resolution with Progressive Vehicle Search and Domain Priori GAN

Wu Liu, Xinchen Liu, Huadong Ma, Peng Cheng
Beijing Key Laboratory of Intelligent Telecommunications Software and Multimedia,
Beijing University of Posts and Telecommunications, Beijing 10086, China
{liuwu,xinchenliu,mhd,peng_c}@bupt.edu.cn

ABSTRACT

In this paper, we address the challenging problem of vehicle license plate image super-resolution. Different from existing image super-resolution approaches only resorted to one single image, we propose to leverage complementary information from multiple images to recover the license plate numbers. To achieve this goal, we design a principled license plate images super-resolution framework which is composed of two components: progressive vehicle search and Domain Priori GAN (DP-GAN). Particularly, we design a null space based progressive vehicle search approach to retrieve the relevant images captured by different cameras given one vehicle with a low-resolution license plate. To handle the extremely varied license plate images caused by different sensors, times, depths, and viewpoints, we also propose a DP-GAN framework to generate multiple spatial correspondences and high-resolution plate images. In the generator network of DP-GAN, a license plate synthesis pipeline is exploited to generate the nearly canonical license plates. In the discriminator network, a spatial split layer is designed to simultaneously preserve the global and local manufacture standards of the license plate. Finally, a multiple images super-resolution GAN is exploited to combine all the synthetic license plates into one high-resolution image. Different from previous super-resolution criteria mainly focus on pixel-level detail recovery condition, we leverage the downstream tasks, i.e. license plate recognition and vehicle search as criteria. The results on a new collected real-world dataset demonstrate that the proposed method achieves the beyond human-level license plate super-resolution performance for automatic license plate recognition and vehicle search.

CCS CONCEPTS

• **Computing methodologies** → **Reconstruction**; Neural networks; • **Information systems** → *Multimedia and multimodal retrieval*;

MM'17, October 23–27, 2017, Mountain View, CA, USA.
© 2017 ACM. ISBN 978-1-4503-4906-2/17/10. . . $15.00
DOI: http://dx.doi.org/10.1145/3123266.3123422

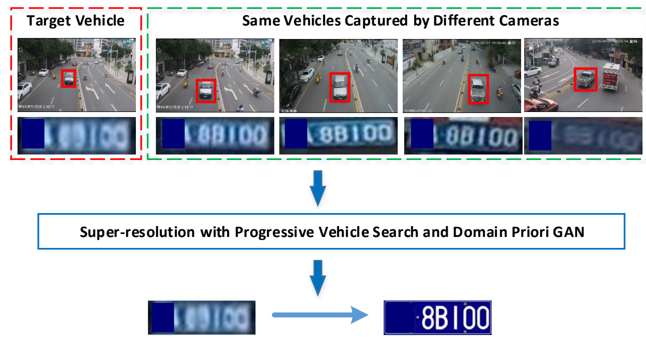

Figure 1: A snapshot of the proposed vehicle search based license plate image super-resolution framework, which exploits the complementary nature among multiple related vehicle images captured by different surveillance cameras to recover the high-resolution license plate image.

KEYWORDS

Super-resolution; Progressive Vehicle Search; License Plate Recognition; Domain Priori GAN; Video Surveillance

1 INTRODUCTION

With the proliferation of vehicles in human life, the vehicles management becomes an important issue in urban surveillance system. As the unique identifier of vehicle, vehicle license plate is always one of the most important research targets in the vehicle detection, recognition, re-identification, and driver behavior modeling. However, due to the complex surveillance background, lighting, camera viewpoints, resolution, motion blur, and occlusion, it is very hard to capture the clear and complete vehicle license plate in urban surveillance. Figure 1 and 2 show some examples, which are randomly selected from images captured by 20 different cameras around a circular road. Even human cannot recognize most of them. Therefore, vehicle license plate super-resolution (SR), which tries to recover and clear the plate images to be better recognized by human and computer, is an important research topic in computer vision and multimedia.

Many existing methods try to enhance the license plate images through SR methods, which can be divided into three categories: 1) the interpolation-based approaches [9], 2) the reconstruction-based approaches [23, 26], and 3) the learning-based approaches [11, 28]. Recently, because of the powerful learning abilities of deep neural networks [3, 6, 16], the data-driven deep learning methods achieve the state-of-the-art

Figure 2: Some results of the proposed multiple license plates super-resolution with progressive vehicle search and domain priori GAN. Even for the plate images which cannot be recognized by humans, our method can well recover their high-resolution images.

performance in single image SR (SSR) [1, 10]. This kind of methods can directly learn a deep neural network to map one low-resolution (LR) image into a high-resolution (HR) one. However, the existing SR methods all try to recover the license plate from one single image. Differently, as shown in Figure 1, in real-world urban surveillance, one vehicle can be captured by more than one camera. Therefore, we can obtain the complementary information from multiple images to recover the license plate numbers.

In this paper, we represent one of the first attempts towards recovering the license plate through multiple images captured by different cameras in the entire city. Given a vehicle image with LR license plate, we use it as a query to search its recorded images captured by different cameras [19]. The next step is how to utilize all these recorded images to recover its real and clear license plate. Admittedly, the multiple images SR (MSR) method can be exploited. However, the existing MSR methods mainly focus on continuous frames in one video, which have same scene with similar occlusions. However, as shown in Figure 1, the license plate images obtained by the vehicle search are extremely varied with different blur, resolution, occlusion, affine transformation, lighting, and viewpoints. These differences bring great challenges to exploit the complementary information among multi-images to recover the final HR image. In addition, we cannot find a fixed transform to super-resolve all of them. Moreover, the vehicle search may bring wrong license plate images from other vehicles. Therefore, as the above challenges, we cannot directly utilize the existing MSR methods to solve this new problem.

To this end, we propose a domain priori generative adversarial networks (DP-GAN) to align and recover the license plate images from different sensors, times, depths, and viewpoints into spatial corresponding and HR images. The DP-GAN contains two adversarial networks — Generator Network (GN) and Discriminator Network (DN). The GN is a Fully Convolutional Network (FCN) to map the LR license plate images into HR ones. It contains one convolution layer, N residual blocks, and two deconvolution layers. To effectively train GN, we need massive LR-HR license plate images pairs. The LR image can be collected from the real-world surveillance images. However, it is hard to select the HR image as the complicated surveillance environment. We address the issues by render based image synthesis. The canonical HR image is synthesized for each LR according to the strict manufacture standards of license plate. Except clarifying the LR images, all the grotesque LR plate images can be registered into one template for further process. Moreover, the optimization target of traditional FCNs always captures the pixel-level differences between LR-HR image pairs, which neglects the domain priors of vehicle license plates. Hence the GN always generates the irregular characters to significantly decrease the recognition rate. To avoid this phenomenon, we train a VGG network as the DN to differentiate the synthetic plate and canonical HR images. More important, a novel spatial split layer is added into the VGG to preserve the global and local manufacture standards of license plate as domain priors. Finally, the adversarial loss and content loss are tightly intertwined to improve the GN and DN.

Then the proposed DP-GAN is employed in the vehicle search based multiple license plate images SR framework. As shown in Figure 3, we firstly exploit a null-space based progressive vehicle search method to retrieve the recorded images from different cameras. In the coarse filtering, the proposed method embeds the low and high-level appearance attributes into null space to find the appearance similar vehicles. Then a Siamese Neural Network (SNN) is adopted to verify the similarities of license plates for fine search. Finally, the spatio-temporal information is employed to re-rank the results. Next, all the searched license plate images are aligned and super-resolved through the DP-GAN. Then the generated spatial consistent and HR plate images are input into a multi-images SR GAN (MSR-GAN). The MSR-GAN has the similar architecture with DP-GAN, except that the multiple images are set as different channels of input. Through the MSR-GAN, we can get the final HR license plate image for downstream tasks. Some examples can be found in Figure 2. The proposed method can even clearly recover the license plates which cannot be recognized by human. Furthermore, we build a real-world license plate dataset captured by 20 different cameras in a city to evaluate the proposed method. In the evaluation, different from previous SR criteria mainly focus on pixel-level recovery condition, we directly leverage the downstream tasks like license plate recognition and vehicle search as criteria. With the proposed methods, the automatic license plate recognition rate is significantly improved from 7% to 51%, which even outperforms the human-level performance (20% with query only and 49% with vehicle search). Moreover, it also achieves double times improvements (i.e., mAP value from 27.77% to 60.47%) for vehicle search task.

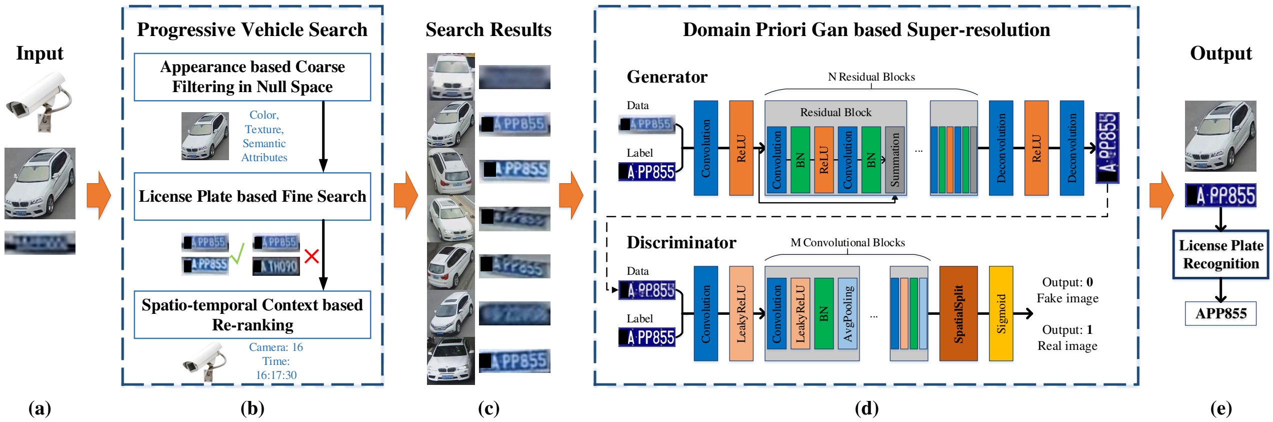

Figure 3: The overview of our proposed multiple license plates super-resolution with progressive vehicle search and domain priori GAN.

In summary, this paper makes the following contributions:

- We have designed an innovative vehicle license plate super-resolution framework that represents one of the first attempts towards utilizing the complementary information among searched multiple license plate images captured by different cameras.
- We propose a DP-GAN to combine prior knowledge from data distribution and domain knowledge of license plate to generate the spatial corresponding and high-resolution plate images.
- We propose a null space based progressive search to retrieve recorded images captured by different cameras, and a multi-images GAN to generate the final high-resolution image.

2 DOMAIN PRIORI GAN

For the general GAN model, two parallel networks GN and DN are trained alternatively to improve each other [4]. GN models the data distribution, and DN estimates the probability that a sample came from the groundtruth rather than GN with domain priori knowledge. Next, the tightly intertwined adversarial training process can significantly improve the performance of GN and DN. For the license plate SR problem, we propose a DP-GAN which contains a GN to take a LR plate I_l as the input to generate a HR plate I_h, and a DN to discriminate whether a HR plate is from the groundtruth I_g or generated by GN. In the training of GAN, the optimization target of GN tries to minimize the pixel-level differences between I_l and I_g. Differently, the DN tries to learn the strict manufacture standards of license plate as domain priors. Moreover, compared with the existing SR-GAN model [12], the innovation of DP-GAN can be summarized: 1) we show that training GAN by massive synthetic data with domain prior is an effective approach for license plate SR; 2) we propose a synthesis pipeline in GN to generate the spatial corresponding and HR license plates for further MSR process; and 3) we design a spatial split layer in DN to focus on both the global and local domain priors of the vehicle license plate. Next, we will introduce these innovations in detail.

2.1 Generator Network with Synthesis Pipeline

The state-of-the-art low-level computer vision tasks, e.g., image super-resolution and deblur, adopt FCNs to learn an end-to-end mapping from the input image to the groundtruth as in [1]. Therefore, we exploit the FCNs as the GN in GAN to learn an end-to-end mapping from I_l to I_h. To improve the capacity of the very deep network and avoid diffusion of gradients during training, as shown in Figure 3 (d), we adopt a hierarchy of N residual blocks proposed in [6] as the backbone of our generator. As in [6], each residual block contains two convolutional layers with 64 filters of 3×3, two batch-normalization layers, and one ReLU activation function in a specific order. A shortcut connects the input and output of the block by pixel-wise summation as the identity mapping to guarantee the gradient flow through the deep network. After the N residual blocks, a deconvolutional layer with stride of two is adopted to enlarge the resolution of the feature map. At the tail of the network, we use a deconvolutional layer with three filters of 1×1 to generate the result three-channel I_h.

Moreover, in generic training of the deep SR networks, the LR images are always obtained from groundtruth HR images with downsampling and affine transformation to generate the LR-HR pairs [10]. However, for license plate SR, we can collect massive LR plate images from real-world surveillance videos. Conversely, it is a big issue to provide the groundtruth I_g for each I_l. Although one solution is to select the highest-resolution plate image from the searched license plate set as the groundtruth, the criterion of the highest resolution is hard to be defined. Terrible groundtruth can significantly decrease the network performance. Inspired by the render for CNN in [22], we propose to address this issue by synthesising the I_g for each I_l. As license plates have strict manufacture standards, it is reasonable and easier to synthesis the groundtruth plate I_g

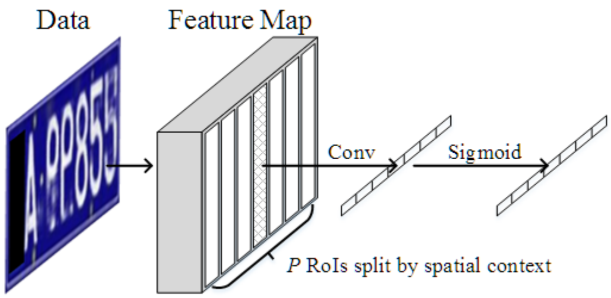

Figure 4: The structure of the proposed spatial split layer in discriminator network.

with these standards. The correct characters of I_l are labeled from the groundtruth. Then the I_g is generated with the characters and license plate standardization. Some examples of I_l and I_g pairs are shown in Figure 7. With the rendered groundtruth, our license plate SR problem can be converted to learn the license plate synthesis process, which combines the image registration and SR in one synthesis pipeline. Moreover, the generated spatial corresponding and HR license plate images can be utilized for further MSR process.

2.2 Discriminator Network with Spatial Split Layer

The DN is trained to differentiate the plate super-resolved by the GN from the groundtruth one. The backbone of the discriminator employs a VGG-like network which contains one convolutional layer and a hierarchy of M convolutional blocks. $M = 3$ in our implementation, the filter numbers in the convolutional layers are 64, 128, and 256 respectively. The LeakyReLU activation function and batch-normalization are used as in [12]. At last, to obtain the probability that an input plate comes from the groundtruth or from GN, a sigmoid function is connected to the tail of the network as the classification task.

The conventional discriminator in [12] takes the input image as a whole to differentiate it came from the generator rather than the groundtruth. However, license plate has strict manufacture standards which can provide valuable spatial context, such as the location, length, width, and structure of each alphabet and number character. To further focus on the local details, as shown in Figure 4, we design a novel Spatial Split Layer (SSL) in the DN. The SSL can make the DN focus on local features for each character and reduce the interference between neighboring ones. Through the original DN, a discriminative feature map can be obtained through the forward propagation of the backbone network. In the SSL, the feature map is split into P Regions of Interest (RoIs) corresponding to the locations of characters in the input plate. For all P RoIs, we then apply a convolutional layer with one filter of $W_{RoI} \times H_{RoI}$ to obtain one scalar for each RoI. Finally, the output of SSL is fed into the sigmoid function with global feature to estimate the probability of the generated license plate. With SSL, the DN can not only consider pixel-level global information of the whole image, but also discriminate local details of each character. During training, the error for each character can be calculated to further train the generator, especially for the characters with complex texture.

2.3 Adversarial Loss Function

As the aforementioned principle in [4], the aim of DP-GAN is to train a generator network G which makes the output I_h indistinguishable from the groundtruth I_g. Therefore, it is critical to define an adversarial loss function for the generator and discriminator.

For the generator network, the adversarial loss function is formulated as:

$$L_G = L_{MSE} + \lambda L_D. \tag{1}$$

It contains two components from GN and DN respectively. The L_{MSE} is the Mean Squared Error (MSE) between the output $G(I_l)$ of GN and the groundtruth I_g, which is defined as:

$$L_{MSE} = \frac{1}{r^2wh}\sum_{x=1}^{rw}\sum_{y=1}^{rh}(I_{g(x,y)} - G(I_l)_{(x,y)})^2, \tag{2}$$

where r is the scale ratio, w and h are width and height of I_g. The L_D is the output of the sigmoid function in the DN, i.e., the probability that $G(I_l)$ is considered as a I_h by DN. As in [12], we formulate L_D as:

$$L_D = \sum_{n-1}^{N} -logD(G(I_l)) \tag{3}$$

where N is the batch size of one iteration.

In each iteration of the training procedure, we firstly feed a pair of I_l and I_g into the generator GN to compute L_{MSE} through forward propagation. After that the discriminator DN takes I_h with $label = 0$ and I_g with $label = 1$ as the input separately to calculate L_D. Then DN is optimized with L_D by backpropagation. At last, L_G is computed by Equation 1, which is used to optimize GN by backpropagation. To this end, the GN and DN are trained in an alternating manner. With the adversarial loss L_D, the GN can not only make the super-resolved plate approximate the groundtruth by L_{MSE}, but also utilize DN with L_D to prevent GN generating malformed characters in the plate.

3 PROGRESSIVE VEHICLE SEARCH BASED LICENSE PLATES SR

As shown in Figure 3 (b), given the vehicle image whose license plate need to be recognized, we employ a null space based progressive vehicle search process to obtain its related images captured by other surveillance cameras. The proposed vehicle search method is inspired by [18] in Internet of Thing Search [19].

Appearance Attributes based Coarse Filtering in Null Space. In the real-world practice, the vehicles' appearance features are efficiently to be extracted and effectively to filter out the most dissimilar vehicles in the large-scale traffic surveillance dataset. Therefore, we extract the vehicles' low and high-level visual appearance as coarse filtering. The low-level features contain Bag-of-visual-Words (BoW) as texture feature and Color Name Model (CNM) as the color feature. Moreover, we extract the high-level attribute features with deep CNN, i.e., the GoogLeNet trained on the CompCars dataset [30], to capture rich semantic attributes, such as the shape of lights, the number of doors, and type of cars, which are very important for vehicle search. Different from [18], we adopt a kernel based Null Foley-Sammon Transform (NF-ST) metric learning approach [22] to combine the low and high-level features in null space to identify the same vehicles in different camera views. In this null space, images of the same vehicles are collapsed into a single point with rigorous theoretical proof on the resulting subspace dimension. The kernel based NFST can deal with the non-linearity of the vehicle's appearance, which further boosts the matching performance within the null space. Therefore, the combination of multi-level appearance features in the null space offers a perfect solution to the challenging vehicle search problem.

License Plate Verification based Fine Search. After the search space narrows down from the whole vehicle database to a relatively small one after coarse filtering, we utilize the license plate to further find the same vehicle license plates. However, directly recognizing the LR license plate image is nearly impossible. Therefore, we train a SNN to calculate the license plates similarities between query and source vehicles [18]. In the implementation, the SNN designed for plate verification contains two parallel CNNs connected with a the contrastive loss layer. During training, the similar or dissimilar pairwise plate images are fed into the two parallel CNNs separately. In the test stage, we use the learned SNN to extract the feature of the last fully connected layer from the plate images. The Cosine distance is adopted to estimate the similarity scores of the two plates.

Spatiotemporal Property based Re-ranking. In real-world practice, it is reasonable to perform the vehicle search with a from-near-to-distant fashion in both the spatial and temporal scales. Based on this principle, we exploit the spatiotemporal property proposed in [18] to further re-rank the vehicle Re-Id results.

Multiple license plate images based generative adversarial network (MSR-GAN). With the above progressive vehicle search method, we can retrieve a set of same license plate images recorded by different cameras for the query images. However, as we introduced before, the searched images are extremely different from each other because of the different sensors, times, depths, and viewpoints. These differences seriously affect the performance of the MSR method. Therefore, we firstly feed these images into the DP-GAN respectively to get the spatial corresponding and HR license plate images. Then a MSR-GAN is exploited to generate the final HR image. The MSR-GAN has the similar GN and DN architectures with DP-GAN. The main difference is that each aligned license plate image is set as one channel of the input image. We select the top nine similar vehicles as nine different channels. The MSR-GAN is retrained as the similar strategies of DP-GAN. Finally, we get one final HR license plate image from MSR-GAN for further process.

4 EVALUATIONS

4.1 Evaluation Dataset

To facilitate related research and well evaluate the proposed method, we collect and label all the vehicle license plates in the VeRi dataset [18]. VeRi has 776 different vehicles and 9000 records captured by 20 surveillance cameras installed along several roads in an 1.0 km^2 area, which guarantees data quality and real-world traffic scenarios. The cameras are installed in arbitrary positions and directions. Each vehicle is captured by at least two cameras in varied viewpoints, lightning conditions, and backgrounds, which guarantees practical urban traffic environment. Then ten volunteers are asked to label all the vehicle plate numbers. We select 746 vehicles and $24,349$ plate images, where each vehicle has at least one license plate image can be recognized by human as groundtruth. Then the groundtruth images are removed. Finally, we randomly select 594 vehicles ($19,524$ plate images) as training and other 152 vehicles ($4,825$ plate images) as testing set. For each vehicle, one LR image is randomly selected as query image. For each vehicle in training, we generate one groundtruth license plate image with strict rules for training [1].

4.2 Experimental Setting

The traditional super-resolution problems always choose mean squared error (MSE) and peak signal-to-noise ratio (PSNR) as the evaluation criteria. However, as we introduced before, different from focusing on pixel-wise image differences, the goal of license plate super-resolution is making human or computer correctly recognize the license plate number. So in our evaluation, we choose two new evaluation criteria: the recognition rate of human and automatic license plate recognition software. For human evaluation, we invite 12 volunteers to recognize the plate number. For automatic license plate recognition, we choose an open source application named "EasyPR"[2]. To protect the privacy, the VeRi covers the left 3 characters. Hence the recognition rate is computed according to the right four characters.

4.3 Evaluation on Super-resolution Methods

First of all, we compare the following seven different super-resolution methods on the proposed license plates dataset to recover searched multiple LR license plate images:

(1) **Baseline**. This is a baseline method without any SR.
(2) **VDSR [10]**. This an end-to-end network with 20 convolutional layers for SSR [10]. According to their evaluation, the performance of VDSR is better than other state-of-the-art single image SR methods. We retrain the network

[1] The dataset is released at "https://github.com/dpgan/mlpsr" as a benchmark for vehicle license plates SR in this community.
[2] "EasyPR," https://github.com/liuruoze/EasyPR

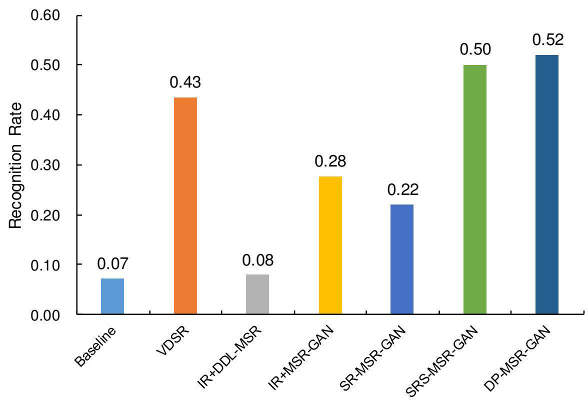

Figure 5: Evaluation on different super-resolution methods with automatic license plate recognition.

on our training set with rendered images and exploit the late fusion to combine the multiple license plates.

(3) **IR+DDL-MSR [14]**. We firstly employ the fast predictive image registration method in [31] to align all the plate images into spatial corresponding. Then we directly use the deep draft-ensemble learning based multiple images SR method proposed in [14] to super-resolve the multiple LR images into one HR image.
(4) **IR+MSR-GAN**. The multiple LR images are also aligned by the fast predictive image registration method in [31]. Then we directly train a MSR-GAN with these aligned images for SR.
(5) **SR-MSR-GAN [12]**. We directly train the original SR-GAN [12] without synthesis pipeline. The license plate images that can be accurately recognized by EasyPR are set as I_g.
(6) **SRS-MSR-GAN**. Different from the above scheme, we retrain the original SR-GAN [12] with the rendered I_g.
(7) **DP-MSR-GAN**. The proposed DP-GAN based MSR method.

The results are shown in Figure 5. First of all, all the SR based methods can improve the automatic performance compared with baseline. Moreover, the results of IR+DDL-MSR and IR+MSR-GAN show that, although we try to align the plates with diverse viewpoints and affine transformations into spatial corresponding, the state-of-the-art MSR methods cannot well recover the images. The main reason is that the existing MSR methods cannot effectively process and fuse the diverse LR images which have extremely large differences among them. Differently, through DP-GAN, the recovered multiple license plate images are spatial corresponding and clearer. Therefore, the DP-GAN can help the MSR method more sufficiently exploit the complementary information to improve the SR performance. The comparison between SR-MSR-GAN and SRS-MSR-GAN shows that training GAN by massive synthetic data with domain prior is an effective approach for license plate SR. The comparison between VDSR and SRS-MSR-GAN demonstrates the positive effects of the DN and adversarial loss function. Finally, the proposed DP-MSR-GAN achieves the best performance, which

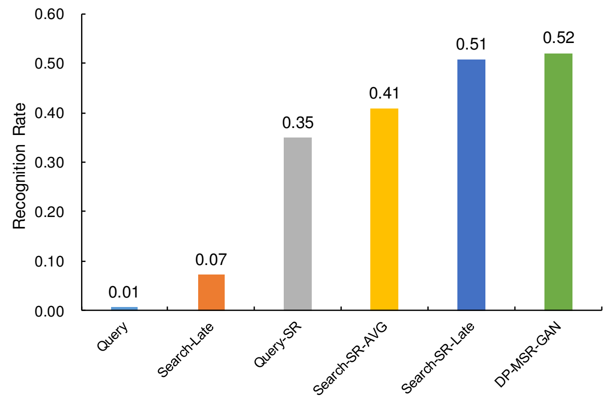

Figure 6: Evaluation on progressive vehicle search multi-images super-resolution methods with automatic license plate recognition.

demonstrates that the proposed synthesis pipeline and SS-L can more effectively combine prior knowledge from data distribution and local domain knowledge to improve the SR performance.

4.4 Evaluation on Vehicle Search Based Super-resolution

To demonstrate the effectiveness of utilizing the progressive vehicle search based multiple license plates SR, we compare the results with and without the vehicle search based methods. For search based method, we select the top eight similar vehicle images. The proposed DP-GAN is chosen as the SR method.

(1) **Query**. Only using the original LR query image without SR for automatic recognition.
(2) **Search-Late**. Directly using the EasyPR to recognize the original LR query and each searched LR image without SR. The nine recognition results are combine with late fusion strategies.
(3) **Query-SR**. Only using the HR query image recovered by DP-GAN for automatic recognition.
(4) **Search-SR-AVG**. We combine the HR query and searched images super-resolved by DP-GAN into one HR image with average fusion for automatic recognition.
(5) **Search-SR-Late**. We combine the automatic recognition results of the HR query and searched images super-resolved by DP-GAN with late fusion strategies.
(6) **DP-MSR-GAN**. The HR query and searched images super-resolved by DP-GAN are input into the MSR-GAN to generate one HR image. The the HR image is used for automatic recognition.

The results are shown in Figure 6. From the comparison between Query and Search-Late, we can find that regardless of whether using the SR method or not, the progressive search method can significantly improve the SR performance. It demonstrates that the complementary information among searched license plate images is definitely useful for vehicle plate super-resolution. Moreover, through comparing the Query and Query-SR, Search-Late and Search-SR-Late, we

Table 1: Subjective evaluation on vehicle Search based super-resolution with human recognition.

Methods	Q	Q-SR	QS	QS-SR	**DP-MSR-GAN**
RR	0.20	0.39	0.49	0.52	**0.54**

Table 2: Evaluation of the proposed super-resolution method with vehicle search.

methods	mAP(%)	HIT@1(%)	HIT@5(%)
FACT [17]	18.49	50.95	73.84
PROVID [18]	27.77	61.44	78.78
PVS	53.29	81.76	94.70
PVS + DP-GAN	**60.47**	**85.52**	**95.11**

can find that the proposed DP-GAN can significantly improve the recognition performance on only using query and vehicle search based method respectively. Moreover, we also compare different multiple image super-resolution strategies. Compared with average and late fusion, the proposed DP-MSR-GAN strategy achieves the best performance, which demonstrates that the proposed multiple super-resolution network can more effectively mine the complementary information among the researched related images. Moreover, besides the higher recognition rate, the DP-MSR-GAN can combine the multiple LR images into one HR image, which is more convenient for further process. Moreover, on the VeRi dataset, the progressive search process needs $32.5ms/query$, DP-GAN needs $4ms/plate$, and MSR-GAN needs $13ms$ on a server with Xeon E5-2660v3 2.6GHz CPU, four NVIDIA TESLA K80 GPU cards, 256G DDR4 memory, and 1T SSD disk.

4.5 Subjective Evaluation by Human Recognition

Besides using the EasyPR as an evaluation tool, we also invite 12 volunteers to recognize the license plates in the test set as subjective evaluations. Each volunteer will recognize the license plate with 1) query only (Q), 2) query super-resolved with DP-GAN (Q-SR), 3) query and searched images (QS), 4) query and searched images super-resolved with DP-GAN (QS-SR), and 5) the query and searched images generated with DP-GAN and MSR-GAN (DP-MSR-GAN). The average recognition rates can be found in Table 1. From the results, we can find that the proposed DP-GAN can significantly improve the human recognition rate on single license plate image from 0.20 to 0.39. Moreover, with progressive searched images, the human recognition can be improved to 0.49.

Moreover, QS-SR is also better than QS. The above results demonstrate that the proposed DP-GAN achieves beyond human-level license plate SR performance. Through comparing the QS-SR and DP-MSR-GAN, we can find that the latter one not only achieves better performance, but also significantly improves human recognition efficiency. As for search-SR, the volunteers have to see nine images and artificially combine the recognition results for one license plate. However, for DP-MSR-GAN, the volunteers only need to see one HR image. Through comparing the automatic (Search-MSR in Figure 6) and human (QS in Table 1) recognition results, we can also find that the proposed method outperforms the human recognition performance (0.51 VS. 0.49). It also demonstrates the beyond human-level license plate SR performance.

4.6 Evaluation by Vehicle Search Task

In our framework, vehicle search is the prerequisite to obtain multiple plate images for one query. First of all, we compare different vehicle search methods, i.e., FACT [17], PROVID [18], and our null space progressive search (PVS) on VeRi dataset. The mean Average Precision (mAP), HIT@1, and HIT@5 are shown in Table 2. From the results, we can find the PVS achieves the best performance. Moreover, in return, we will discuss whether the proposed DP-GAN can further improve progressive vehicle search. For the second procedure of progressive vehicle search framework, i.e., license plate verification based fine search, we adopt the DP-GAN as a preprocessor for all query, source, and training license plate images. With the super-resolved plate images, we retrain the SNN for plate verification. From the results, we can find that after adding the DP-GAN into the license plate verification, the performance is greatly improved again. The results further demonstrate the effectiveness of the proposed license plate super-resolution method [3].

4.7 Qualitative Results

The examples of different license plate super-resolution methods can be found in Figure 7. For the two query only based results, i.e., Q-VDSR and Q-DP-GAN, we can find that there are more strange characters in results of Q-VDSR. Moreover, compared with IR+MSR-GAN, the power of DP-GAN also help the DP-MSR-GAN achieve much better performance. The results demonstrate the DP-GAN can help the GN generate more accurate license plates with strict global and local manufacture standards. Moreover, with the synthetic pipeline in DP-GAN, we can get the spatial corresponding and clear multiple license plate images, which can make MSR-GAN better exploit the complementary information among the multiple license plates captured by different cameras. Undoubtedly, the proposed License Plate Super-resolution with Progressive Vehicle Search and Domain Priori GAN achieves the best performance. Finally, the below two lines give the failure examples of all methods. The main reasons are that the "0" and "Q", "8" and "B" are too similar to discriminate each other. The same situations are also happened on other characters. According to our statistics, the "P, T, A, 2, J" have the highest recognition rates for both software and human. Conversely, the "Y, D, G, B, Q" are the lowest ones.

5 RELATED WORK

5.1 Single-Image Super-Resolution

Single Image Super-Resolution (SSR), which is to generate a HR image from a LR image, has attracted numerous researchers due to its great importance for computer vision

[3]More evaluations of the vehicle search methods and qualitative searched results are in "https://github.com/dpgan/mlpsr".

Figure 7: The examples of different license plate super-resolution methods.

and multimedia tasks [27]. The generic SSR approaches can be categorized into three types: 1) the interpolation-based approaches, 2) the reconstruction-based approaches, and 3) the learning-based approaches.

In the early years, interpolation-based algorithms, such as bilinear, bicubic, and Lanczos, have been proposed as basic approaches to generate HR images from LR images through pre-defined computational formulas [9]. Although these algorithms are very efficient due to the pixel-level processing, they may generate artifacts and blur on the results, which cannot satisfy the HR requirements in real practice [32]. The reconstruction-based algorithms usually generate HR images by adopting texture-level priors, such as the edge prior [26] and the gradient profile prior [23]. These priors are usually obtained from a large number of natural images by statistics and utilized as a constraint when generating a HR image. While this type of algorithms can guarantee sharp edges with limited artifacts, they neglect local structures or complex high-frequency details.

In recent years, the learning-based algorithm, also known as the example based algorithm, has attracted more researches for SSR. These methods adopt machine learning algorithms, such as sparse coding [29], dictionary learning [28], and regression [11], to learning a LR-HR mapping from large-scale training patches. Moreover, CNNs have also been explored on SSR task by directly learning a mapping network from a large number of LR-HR image pairs and outperform the conventional methods [1, 10, 21]. Recently, Ledig *et al.* [12] employ a SR-GAN with perceptual loss to learn a generative model for HR image generation. They alternately train two deep CNNs: CN maps the LR images to HR ones, and a DN differentiates the super-resolved HR images and original photo-realistic images. The SR-GAN has achieved better visual effect than existing methods on the proposed mean-opinion-score metric. In addition, many domain specific SSR methods have been proposed to generate HR images of faces [5, 7], vehicle license plates [25], and scenes [24]. These methods usually integrate domain priors with existing SSR algorithms and achieve excellent performance in specific domains. The above related works inspire us to design a DP-GAN, which combines a adversarial network integrated with domain priors to improve the vehicle search-based license plate super-resolution task.

5.2 Multi-Image Super-Resolution

Our vehicle search-based license plate SR framework is basically a MSR problem. Existing MSR approaches have usually focused on multi-frame or video super-resolution [2, 8, 13, 15, 20]. Early works have adopted basic models of SSR to the MSR problem but neglected the temporal information in multiple frames or video sequences [2]. Differently, Liu *et al.* [15] propose a Bayesian approach to adaptive video super-resolution by estimating motion, blur and noise parameters simultaneously while generating the HR frames from the LR frames. To handle the motion blur in MSR, Ma *et al.* [20] propose an EM framework to guide residual blur estimation and HR frame reconstruction. The CNNs and Recurrent Neural Networks (RNNs) have been also adopted to simultaneously learn the spatial mapping and temporal relations in video sequences and achieved the state-of-the-art results [8]. However, in our problem, the license plates captured by multiple surveillance cameras are not sequential frames and lack the temporal information for Bayesian or motion estimation. Moreover, the existing methods cannot effectively process the extremely varied license plate images from different sensors, times, depths, and viewpoints

6 CONCLUSION

In this paper, we have investigated the issue of license plate SR with diverse plate images of one vehicle captured by different cameras. First of all, a null space based progressive vehicle search process is exploited to retrieve these recorded images. Furthermore, a DP-GAN is proposed to align and recover these diverse images into corresponding and high-resolution. In the DP-GAN, we propose a license plate synthesis pipeline with domain prior in the generator network. Moreover, a spatial split layer is designed to restrict the global and local rules of license plate in the discriminator network. The adversarial loss function in DP-GAN can combine the two restrictions to improve each other simultaneously. Finally, all the recovered images are fed into a MSR-GAN to generate the final HR image. The evaluations not only demonstrate the proposed method can significantly beyond the state-of-the-art single and multiple images SR methods, but also make the automatic license plate recognition software achieves human-level performance. Moreover, we also demonstrate that the proposed license plate SR method can significantly improve the vehicle search in return.

7 ACKNOWLEDGEMENT

This work is partially supported by the National Key Research and Development Plan (No. 2016YFC0801005), the Funds for Creative Research Groups of China (No. 61421061), the National Natural Science Foundation of China (No. 61602049), and the Beijing Training Project for the Leading Talents in S&T (No. ljrc201502).

REFERENCES

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2016. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 38, 2 (2016), 295–307.

[2] Sina Farsiu, M Dirk Robinson, Michael Elad, and Peyman Milanfar. 2004. Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing* 13, 10 (2004), 1327–1344.

[3] Chuang Gan, Naiyan Wang, Yi Yang, Dit-Yan Yeung, and Alexander G. Hauptmann. 2015. DevNet: A Deep Event Network for multimedia event detection and evidence recounting. In *IEEE CVPR*. 2568–2577.

[4] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *NIPS*. 2672–2680.

[5] Zhen Han, Junjun Jiang, Ruimin Hu, Tao Lu, and Kebin Huang. 2012. Face image super-resolution via nearest feature line. In *ACM Multimedia*. 769–772.

[6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *IEEE CVPR*. 770–778.

[7] Junjun Jiang, Ruimin Hu, Zhongyuan Wang, and Zhen Han. 2014. Face super-resolution via multilayer locality-constrained iterative neighbor embedding and intermediate dictionary learning. *IEEE Transactions on Image Processing* 23, 10 (2014), 4220–4231.

[8] Armin Kappeler, Seunghwan Yoo, Qiqin Dai, and Aggelos K Katsaggelos. 2016. Video super-resolution with convolutional neural networks. *IEEE Transactions on Computational Imaging* 2, 2 (2016), 109–122.

[9] Robert Keys. 1981. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 29, 6 (1981), 1153–1160.

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

[11] Kwang In Kim and Younghee Kwon. 2010. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 6 (2010), 1127–1133.

[12] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and others. 2016. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint arXiv:1609.04802* (2016).

[13] Yawei Li, Xiaofeng Li, Zhizhong Fu, and Wenli Zhong. 2016. Multiview Video Super-Resolution via Information Extraction and Merging. In *ACM Multimedia*. 446–450.

[14] Renjie Liao, Xin Tao, Ruiyu Li, Ziyang Ma, and Jiaya Jia. 2015. Video super-resolution via deep draft-ensemble learning. In *IEEE ICCV*. 531–539.

[15] Ce Liu and Deqing Sun. 2014. On Bayesian adaptive video super resolution. *IEEE Transactions on Pattern Analysis And Machine Intelligence* 36, 2 (2014), 346–360.

[16] Wu Liu, Tao Mei, Yongdong Zhang, Cherry Che, and Jiebo Luo. 2015. Multi-task deep visual-semantic embedding for video thumbnail selection. In *IEEE CVPR*. 3707–3715.

[17] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. 2016. Large-scale vehicle re-identification in urban surveillance videos. In *IEEE ICME*. 1–6.

[18] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. 2016. A Deep Learning-Based Approach to Progressive Vehicle Re-identification for Urban Surveillance. In *ECCV*. 869–884.

[19] Huadong Ma and Wu Liu. 2017. Progressive Search Paradigm for Internet of Things. *IEEE MultiMedia* PP, 99 (2017), 1–1. DOI: https://doi.org/10.1109/MMUL.2017.265091429

[20] Ziyang Ma, Renjie Liao, Xin Tao, Li Xu, Jiaya Jia, and Enhua Wu. 2015. Handling motion blur in multi-frame super-resolution. In *IEEE CVPR*. 5224–5232.

[21] Haiyi Mao, Yue Wu, Jun Li, and Yun Fu. 2016. Super Resolution of the Partial Pixelated Images With Deep Convolutional Neural Network. In *ACM Multimedia*. 322–326.

[22] Hao Su, Charles Ruizhongtai Qi, Yangyan Li, and Leonidas J. Guibas. 2015. Render for CNN: Viewpoint Estimation in Images Using CNNs Trained with Rendered 3D Model Views. In *IEEE ICCV*. 2686–2694.

[23] Jian Sun, Zongben Xu, and Heung-Yeung Shum. 2008. Image super-resolution using gradient profile prior. In *IEEE CVPR*. 1–8.

[24] Libin Sun and James Hays. 2012. Super-resolution from internet-scale scene matching. In *IEEE ICCP*. 1–12.

[25] Kaggere V Suresh, G Mahesh Kumar, and AN Rajagopalan. 2007. Superresolution of license plates in real traffic videos. *IEEE Transactions on Intelligent Transportation Systems* 8, 2 (2007), 321–331.

[26] Yu-Wing Tai, Shuaicheng Liu, Michael S Brown, and Stephen Lin. 2010. Super resolution using edge prior and single image detail synthesis. In *IEEE CVPR*. 2400–2407.

[27] Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang. 2014. Single-image super-resolution: A benchmark. In *ECCV*. 372–386.

[28] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Thomas Huang. 2012. Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing* 21, 8 (2012), 3467–3478.

[29] Jianchao Yang, John Wright, Thomas S Huang, and Yi Ma. 2010. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing* 19, 11 (2010), 2861–2873.

[30] Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. 2015. A Large-Scale Car Dataset for Fine-Grained Categorization and Verification. In *IEEE CVPR*. 3973–3981.

[31] Xiao Yang, Roland Kwitt, and Marc Niethammer. 2017. Quicksilver: Fast Predictive Image Registration - a Deep Learning Approach. *CoRR* abs/1703.10908 (2017).

[32] Kaibing Zhang, Dacheng Tao, Xinbo Gao, Xuelong Li, and Zenggang Xiong. 2015. Learning multiple linear mappings for efficient single image super-resolution. *IEEE Transactions on Image Processing* 24, 3 (2015), 846–861.