SRR-GAN: Super-Resolution based Recognition with GAN for Low-Resolved Text Images

Ming-Chao Xu[1, 2], Fei Yin[1, 2] and Cheng-Lin Liu[1, 2, 3]
[1]National Laboratory of Pattern Recognition, Institute of Automation of Chinese Academy of Sciences,
95 Zhongguancun East Road, Beijing 100190, P.R. China
[2]School of Artificial Intelligence, University of Chinese Academy of Sciences, Beijing 100049, P.R. China
[3]CAS Center for Excellence of Brain Science and Intelligence Technology, Beijing 100190, P.R. China
Email: xumingchao2017@ia.ac.cn, {fyin, liucl}@nlpr.ia.ac.cn

***Abstract*—Text images convey important information for various applications, while the recognition of low-resolution text images is a challenge. Most existing methods solve this problem using a cascaded scheme in two steps: image super-resolution and high-resolution text recognition. In this paper, we propose a novel framework, called SRR-GAN, which integrates text recognition with super-resolution via adversarial learning. By joint training of recognition and super-resolution models, more generic features for images of various quality can be learned, so as to yield high recognition performance for both high-resolution and low-resolution images. Experiments on natural scene and handwritten texts demonstrate that SRR-GAN outperforms the cascaded scheme on low-resolution images. The results show that SRR-GAN can improve recognition accuracies by 10%-20% relatively on five datasets of scene/handwritten texts. Meanwhile, SRR-GAN maintains high performance on high-resolution images.**

***Index Terms*—Super-Resolution, Adversarial Learning, Text Recognition**

I. Introduction

Text recognition has been studied intensively in past decades. Many methods have been proposed for handwritten text recognition [1] and scene text recognition [2] [3]. The main difficulty of handwriting recognition lies in the large deformation and variation of writing styles. While in scene text recognition, the clutter background and illumination variability of text image make recognition difficult. Text recognition methods can be roughly categorized into segmentation-based methods and segmentation-free methods. In recent years, deep learning techniques have led to large improvement of performance in text recognition. Nevertheless, the recognition of low-resolution text images has received insufficient attention. Super-resolution techniques show potential of improving the performance of low-resolution text recognition.

Previous methods of low-resolution text recognition [4] [5] [6] usually operate in two stages: restore image using super-resolution and then recognize on restored high-resolution image. Deep learning based super-resolution methods [7] [8] [9] have been proposed to generate high-resolution images with higher performance than traditional methods. These methods are applicable to text recognition for improving the quality of text image. However, they have a disadvantage that the correlation between super-resolution and recognition is ignored. In fact, images with different resolutions show different distributions in the feature space, as illustrated in Fig. 1. If the super-resolution model is designed separately from the text recognizer, the restored high-resolution image does not necessarily generate feature distributions suitable for recognition.

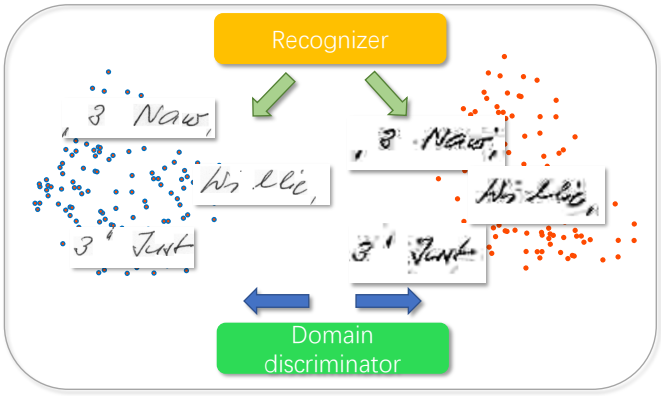


Fig. 1. Different resolution data have different distributions in the feature space.

In this paper, we propose a novel framework, named SRR-GAN, for low-resolution text recognition by optimizing super-resolution model and recognizer jointly. SRR-GAN is composed of a generator, a discriminator and a recognizer. The generator maps low-resolution image input to super-resolution image, the discriminator judges the genuineness of generated high-resolution image, and the recognizer operates on the restored image to give recognition result while using a shared convolutional feature extraction network with the discriminator. Our experiments on public datasets demonstrate that the proposed method can improve the recognition performance on low-resolution scene text and handwritten text images, and outperforms the cascaded scheme of separate super-resolution and recognition. Particularly, the SRR-GAN method performs well on different inputs of high-resolution, low-resolution, and interpolation restored images.

The rest of this paper is organized as follows. Section II briefly reviews related works; Section III describes the

proposed method; Section IV presents experimental results, and Section V concludes the paper.

II. Related Work

Some previous works related to general and low-resolution text recognition are reviewed below.

A. Text Recognition

Recognition of text images, either scene texts (mostly printed texts on cluttered background) or handwritten texts (in paper documents), has been pursued for several decades. The numerous methods proposed so far can be grouped in two rough categories: segmentation-based methods and segmentation-free methods. In the former category (such as [10] [11]), candidate characters detected in text detection stage or generated in an over-segmentation stage are classified by a character classifier, and the classification results are fused, possibly with contexts, to infer the character label sequence.

In the category of segmentation-free methods, Hidden Markov Model (HMM) based methods and recurrent neural network (RNN) based methods are typical examples. Especially, RNN based methods got dominant in recent years. The convolutional RNN (CRNN) method [3] with CTC decoder has attracted high interests. In some works [12] [13] [14], attention networks are used for decoding for extending to recognition of curved text lines.

Some methods [15] [2] use a classifier to classify sliding windows over the text line, and based on the classification results, use a decoder (such as CTC) to infer the character sequence. This can be viewed as an intermediate between segmentation-based and segmentation-free methods.

Taking advantage of deep learning with large amount of data, all these recognition methods have achieved high performance. However, recognition on low-resolution images has not received high attention.

B. Two-stage Low-resolution Text Recognition

For low-resolution text recognition, most previous methods first restore the image to high resolution and then use a recognizer on the restored image.

Super-resolution (SR) has been intensively studied in computer vision to generate high-resolution restored image. In recent years, deep learning based methods [16] [7] [17] [18] [8] have been proposed and shown superior performance. Some researchers have applied super-resolution to text recognition for improving the recognition performance. For instance, Wang et al. [4] utilized an improved Conditional Generative Adversarial Network to conduct text image super resolution. Then based on the SR images, an open source OCR engine (Tesseract) was used to get the recognition result. Zhang et al. [5] concerned OCR performance more than SR performance, and proposed a new loss function to improve the performance of CNN-based text image SR methods. In the end, they used the open Tesseract-OCR software for recognition.

These methods have been shown effective to improve the recognition performance on low-resolution images. However, the super-resolution model and text recognition are designed separately. This can not yield optimal recognition performance because the super-solution model is not designed to directly fit recognition.

III. Proposed Method

A. Overall Architecture

An overview of our framework is shown in Fig. 2. SRR-GAN is composed of three components networks: a super-resolution generator (G), a discriminator (D) and a recognition model (R). For a high-resolution image I_{HR} of size $W \times H \times C$ with C color channels, we get its low-resolution version I_{LR} of size $rW \times rH \times C$ by downsampling from I_{HR} with the factor r ($r < 1$). First, I_{LR} is transformed to super-resolution image I_{SR} through G. Then I_{SR} is fed to D and R, which share the feature extractor layer. In our experiment, we choose convolutional RNN (CRNN) [3] as the recognition network. The features extracted by the shared backbone CNN of D and R are fed into the following discriminator and recognizer. G, D, and R are jointly trained by adversarial learning to optimize the super-resolution and recognition performance. In training, the CNN feature extractor is aimed to extract resolution-independent features under the objectives of D and R, such that accurate recognition can be achieved on images of various resolutions.

B. Formulation

On an input image I, denote the recognition loss as $L_r(I;\theta_r)$, where θ_r denotes the parameters of R. In this work, we use the CTC loss [19] for text recognition. Given the output sequence l and label y, the recognition loss is expressed as

$$L_r(I;\theta_r) = -E_{I\sim p_{train}}[p(l|y;\theta_r)]. \tag{1}$$

The discriminator D is used to judge whether the image is generated or not. We denote the label as y_d, with $y_d = 0$ signifying generated super-resolution images and $y_d = 1$ for original high-resolution image. So, the discriminator loss is expressed as

$$\begin{aligned} L_d(I_{SR}, I_{HR};\theta_d) = &-E_{I_{SR}\sim p_{G(I_{LR})}}[p(y_d = 0|I_{SR};\theta_d)] \\ &-E_{I_{HR}\sim p_{train}}[p(y_d = 1|I_{HR};\theta_d)], \end{aligned} \tag{2}$$

where θ_d denotes the parameters of D, I_{SR} is the generated super-resolution image, and I_{HR} is the original high-resolution image.

G in SRR-GAN is learned to generate super-resolution images that can fool D. To this end, we model L_g using three parts of loss: MSE loss L_{mse}, total variation loss L_{tv} [20] and adversarial loss L_{gen}. The formulation is expressed as

$$L_{mse} = \frac{1}{2}||I_{SR} - I_{HR}||^2, \tag{3}$$

$$L_{tv} = E_{i,j,k}\left[\sqrt{(\hat{I}_{i,j+1,k} - \hat{I}_{i,j,k})^2 + (\hat{I}_{i+1,j,k} - \hat{I}_{i,j,k})^2}\right], \tag{4}$$

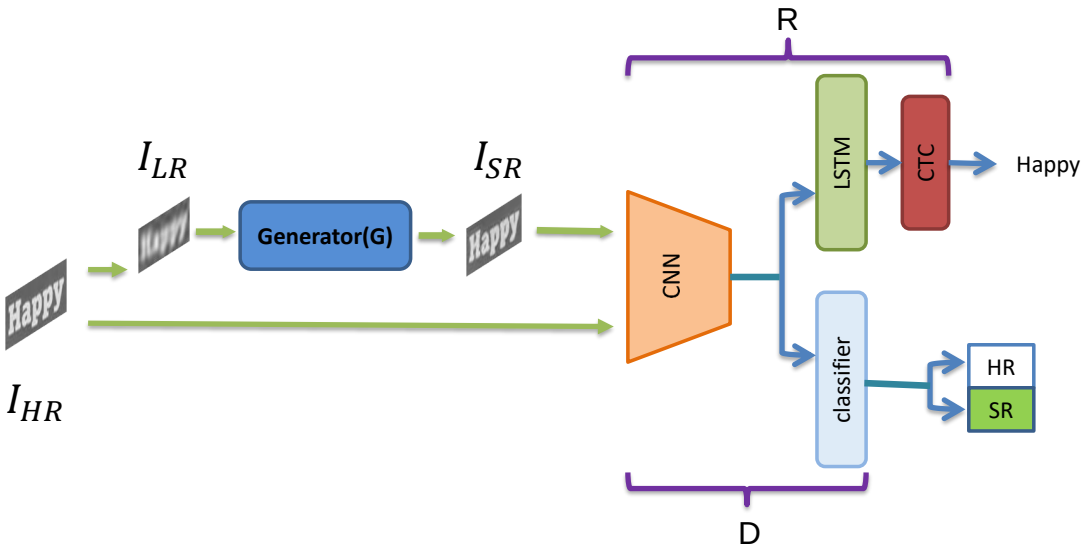


Fig. 2. SRR-GAN is composed of three parts: recognition model (R), discriminator (D), generator (G). D and R share the feature extraction layers.

$$L_{gen} = \sum_{n=1}^{N} -logD_{\theta_D}(G_{\theta_G}(I_{LR})), \tag{5}$$

where i, j, k are the coordinates of horizontal axis, vertical axis and color channel of image, respectively, N is the number of samples, D_{θ_D} represents the probability that the generated super-resolved image I_{SR} is judged as I_{HR}.

Finally, L_g is formulated as the weighted sum of L_{mse}, L_{tv} and L_{gen}:

$$L_g(I_{LR}, I_{HR}; \theta_g) = L_{mse} + \alpha \cdot L_{tv} + \beta \cdot L_{gen}. \tag{6}$$

For generalization, we just set $\alpha = 2e-8$ and $\beta = 1e-3$ as that in [16] without further finetune.

C. Optimization

In training, SRR-GAN with the objective of Eq. (6), is optimized to distinguish images generated by G from original high-resolution inputs, G is optimized to generate high-resolution images that can fool D, and R is optimized to recognize text images generated by G. The network can be trained from scratch. When D becomes more powerful in distinguishing generated images and original high-resolution inputs, G will strive to generate higher-resolved images to compete with D and R. In other words, D, G and R play the minimax game to solve the following problem:

$$\begin{aligned} &\min_{\theta_g \theta_r} \max_{\theta_d} L(\theta_r, \theta_d, \theta_g) = L_g(I_{LR}, I_{HR}; \theta_g) + \\ &\lambda_1 \cdot L_r(I_{SR}, I_{HR}; \theta_r) - \lambda_2 \cdot L_d(I_{SR}, I_{HR}; \theta_d), \end{aligned} \tag{7}$$

where λ_1 and λ_2 are hyper-parameters. In our experiments, λ_1 and λ_2 are empirically set to 1.

IV. Experiments

We evaluated our proposed method on two main text recognition tasks: scene text recognition and handwriting recognition.

A. Datasets

For scene text recognition task, we use the synthetic dataset (Synth) of [21] as the training dataset. This dataset contains 8 million text images with corresponding ground truth words. And we use four popular benchmarks for performance evaluation: IIIT 5K-Words (IIIT5K) [22] containing 3,000 cropped word test images collected from the Internet, Street View Text (SVT) [23] containing 249 street view images collected from Google Street View, ICDAR 2003 (IC03) [24] containing 251 scene images and ICDAR 2013 (IC13) [25] containing 1,015 cropped word images.

Besides, we evaluated our method in handwriting recognition on the IAM-DB [26], which has 6161 text line images for training and 1861 text line images for testing. In addition, we collected a Low-resolution Database (IAM-LR-DB) corresponding to the images of IAM-DB. This was done by printing the document images of IAM-DB and scanning the pages using a scanner at the lowest scan resolution. We only collected the test set, so as to evaluate the performance without low-resolution training data.

B. Experimental Setup

In our experiments, we choose the CRNN [3] as the baseline recognition network. The CNN module in Fig. 2 is the feature extraction backbone in CRNN, and this feature extraction block is shared by D and R. The discriminator D in Fig. 2 is composed of two convolutional layers with 1×1 filers, followed by LeakyReLU activation function. The last convolution layer of D outputs a sigmoid function to get the prediction probability of the input sample. The architecture of G is the same as the one in SRGAN [16].

The training algorithm was implemented using Pytorch on a NVIDIA GM200 GPU. In order to compare results with the baseline R, we obtain up-sampled high-resolution image I_{LR_interp} by apply bicubic interpolation on I_{LR} to meet

the input size of R. We choose $r = \frac{1}{4}$ as the downsampling/upsampling factor. In scene text recognition, we use the same parameter setting for CRNN as [3]. The text word images are re-scaled to 100×32 for input to the recognizer. For each mini-batch, we randomly choose 16 distinct training images. Optimization is implemented using Adam with $\beta_1 = 0.9$ and a learning rate of 10^{-4}, and training is stopped after 10 epochs. The performance metric for this task is the word recognition accuracy, which is the same in [3].

In handwritten text recognition, we use the same network setting as that for scene text recognition. We found that the most common aspect ratio is about 15 in the training set of IAM-DB, so we resize all training text line images to 1000×72, and training is stopped after 100 epochs. The performance metrics are the character error rate (CER) and word error rate (WER).

C. Recognition Results

We show the recognition results of scene text (word) recognition and handwriting (text line) recognition in Tab. I and Tab. II, respectively. Here, we split the results into three categories, which are based on the input I_{HR}, I_{SR} and I_{LR_interp}, respectively. I_{HR} means the input high-resolution image is fed to the recognizer directly. I_{SR} means the input low-resolution image (obtained by down-sampling the original image) is transformed to super-resolution image, which is fed to the recognizer in SRR-GAN. I_{LR_interp} means the input low-resolution image is transformed to high-resolution by interpolation for recognition. In Table I and Table II, "Two-stage" means super-resolution module and text recognizer are trained separately. In the case of SRR-GAN, the text recognition is trained jointly with the super-resolution module, so, its recognition results on I_{HR} and I_{LR_interp} are also different from those of CRNN and "Two-stage".

In Table I, we can see that by simply up-sampling low-resolution image with interpolation, the recognition accuracy of "Two-stage" I_{LR_interp} is much lower than that of recognition directly on high-resolution image I_{HR}. Using a separate super-resolution module for image restoration in "Two-stage" recognition, the accuracy of I_{SR} is improved compared to I_{LR_interp}, but evidently lower than that of direct recognition on I_{HR}. This indicates though super-resolution can restore low-resolution image, the restoration does not optimize recognition. When optimizing super-resolution and recognizer jointly in SRR-GAN, we can see that the trained recognizer gives similar performance on I_{HR} to CRNN. On I_{SR}, SRR-GAN gives much higher accuracy than "Two-stage" method. This verified the superiority of joint training of super-resolution and text recognizer. This trained recognizer also performs well on restored image I_{LR_interp}, as shown in Table I.

The results of handwriting recognition in Table II show similar observations as Table I. Particularly, the SRR-GAN with joint training of super-resolution and recognizer gives similar performance to CRNN on I_{HR}, outperforms significantly "Two-stage" method on super-resolution restored image and interpolation restored image.

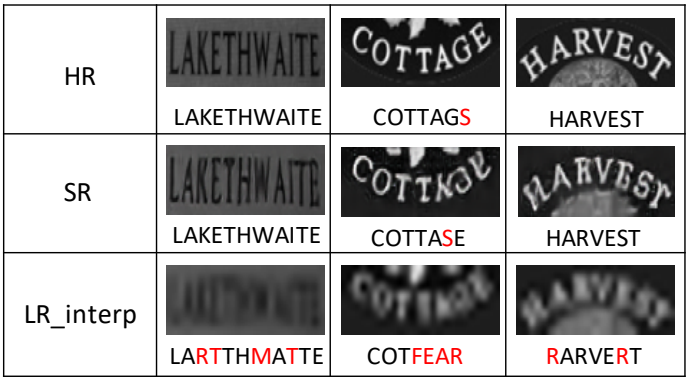


Fig. 3. Some recognition results of our SRR-GAN. It shows that our method can recognize text effectively after getting SR inputs.

Fig. 3 shows some examples of recognition by SRR-GAN on different types of image input. It is shown that by joint training of SRR-GAN, recognition on super-resolution restored image can yield a similar result to recognition on original high-resolution image.

TABLE I
RECOGNITION ACCURACIES(%) OF SCENE TEXT RECOGNITION. "HR", "SR", "LR_interp" DENOTE THE PERFORMANCE OF HIGH RESOLUTION IMAGES, SUPER RESOLUTION IMAGES AND HIGH RESOLUTION RESTORED BY BICUBIC INTERPOLATION.

Acc / Dataset	CRNN [3]	Two-stage		SRR-GAN		
	I_{HR}	I_{SR}	I_{LR_interp}	I_{HR}	I_{SR}	I_{LR_interp}
IIIT5K	78.2	57.9	31.0	**79.1**	**67.4**	**55.5**
SVT	80.8	47.6	22.7	79.6	**56.1**	**43.4**
IC03	89.4	73.2	40.1	**89.4**	**76.8**	**64.3**
IC13	86.7	64.5	37.6	83.1	**67.8**	**59.2**

TABLE II
RESULTS OF HANDWRITING RECOGNITION ON IAM-DB.

IAM	CRNN [3]	Two-stage		SRR-GAN		
	I_{HR}	I_{SR}	I_{LR_interp}	I_{HR}	I_{SR}	I_{LR_interp}
CER	9.9	14.5	51.9	**9.7**	**12.8**	**38.0**
WER	30.0	39.8	80.4	**29.9**	**35.5**	**69.5**

D. Comparison of Adversarial Module and Interpolation

To show the effects of super-resolution on recognition of low-resolution images, we compare the results of different training/testing conditions: 1) Training the recognizer with I_{HR} only; 2) Training the recognizer with interpolation restored image I_{LR_interp} only; 3) Training with both I_{HR} and

TABLE III
RECOGNITION RESULTS OF FOUR DIFFERENT TRAINING CONDITIONS ON IAM-DB.

IAM	Train	I_{HR}		I_{LR_interp}		$I_{LR\_interp} + I_{HR}$		SRR-GAN	
	Test	I_{HR}	I_{LR_interp}	I_{HR}	I_{LR_interp}	I_{HR}	I_{LR_interp}	I_{HR}	I_{SR}
CER		9.9	51.9	21.4	13.1	10.1	14.2	**9.7**	**12.8**
WER		30.0	80.4	50.9	35.9	29.7	37.6	29.9	**35.5**

I_{LR_interp}; 4) Joint training in SRR-GAN. The recognition results on IAM-DB are shown in Table III. We can see that the recognizer training with I_{HR} performs well on high-resolution input image, but poorly on interpolation restored image; The recognizer training with I_{LR_interp} performs well on interpolation restored image, but poorly on high-resolution input; Training with both I_{HR} and I_{LR_interp} gives fairly good performance on both I_{HR} and I_{LR_interp} input images. However, the SRR-GAN gives superior performance on both I_{HR} and I_{LR_interp} input images. This again verifies the necessity and superiority of joint super-resolution and recognizer training.

E. Results of LR Handwriting Recognition

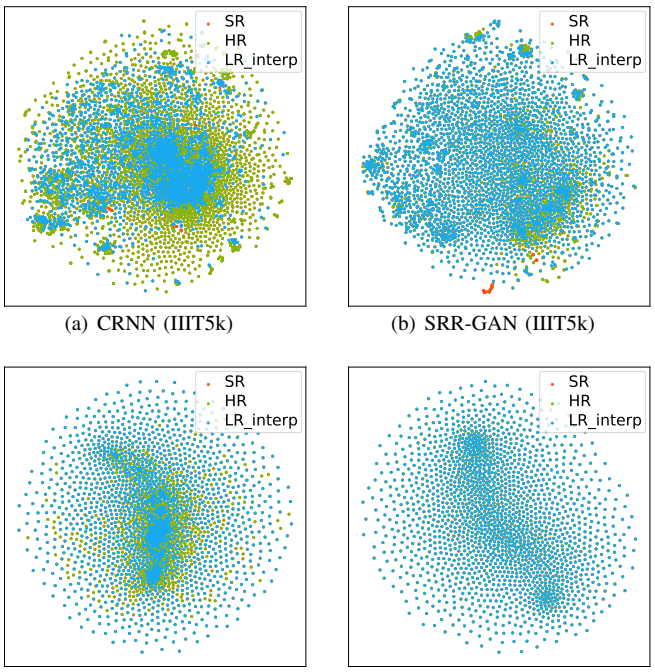
We also tested the recognizer trained in SRR-GAN on our collected low-resolution dataset IAM-LR-DB. The images in IAM-LR-DB were scanned with very low resolution, and so, have much different quality with generated low-resolution data by down-sampling. Some samples are shown in Fig. 1, where we can see that images in IAM-LR-DB are very noisy. In this experiment, the SRR-GAN model is trained with high-resolution and down-sampled images in the training set of IAM-DB, and for testing, our collected noisy low-resolution images are fed into recognizor for recognition directly. From Table IV, we can see that on the scanned low-resolution images, the recognizer CRNN results in high error rates (CER and WER). When using SRR-GAN, the error rates are lower than CRNN, though scanned low-resolution images were not used in training.

TABLE IV
RECOGNITION RESULTS ON IAM-LR-DB.

IAM-LR-DB	CER	WER
SRR-GAN	**59.1**	**92.1**
CRNN	64.5	92.6

F. Visualization of Features

To show how the SRR-GAN with jointly trained super-resolution and recognizer can improve the recognition accuracy, we show the 2-dimensional feature distributions of different types of image input: original high-resolution image (HR), super-resolution restored image (SR) and interpolation restored image (LR_interp). The feature scatter plots of recognizers CRNN and SRR-GAN on three types of input from two datasets are shown in Fig. 4. We can see that the features of HR or SR images by either CRNN or SRR-GAN have good separability, while the features of LR_interp disrtibute compactly with hard separability. By SRR-GAN, the separability of features of LR_interp is largely improved. The improvement of feature separability then leads to increased recognition accuracy.



(a) CRNN (IIIT5k) (b) SRR-GAN (IIIT5k)

(c) CRNN (IAM) (d) SRR-GAN (IAM)

Fig. 4. Feature scatter plots for different types of image input. Left: CRNN; right: SRR-GAN. Green dot: HR; red point: SR; blue: LR_interp.

V. CONCLUSIONS

In this paper, we present a novel framework, called SRR-GAN, for jointly learning super-resolution and recognizer for

low-resolution text recognition. Compared to cascaded (and independently trained) super-resolution and recognition model, the proposed method can make the super-resolution restored image better suit recognition. Experiments results in scene text recognition and handwriting recognition on public datasets demonstrate that the proposed method performs superiorly on low-resolution text images.

VI. Acknowledgments

This work has been supported by the National Natural Science Foundation of China (NSFC) grants 61733007, 61573355 and 61721004.

References

[1] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 5, pp. 855–868, 2009.

[2] F. Yin, Y.-C. Wu, X.-Y. Zhang, and C.-L. Liu, "Scene text recognition with sliding convolutional character models," *arXiv preprint arXiv:1709.01727*, 2017.

[3] B.-G. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 11, pp. 2298–2304, 2016.

[4] Y. Wang, W. Ding, and F. Su, "Super-resolution of text image based on conditional generative adversarial network," in *Pacific Rim Conference on Multimedia*, 2018, pp. 270–281.

[5] H. Zhang, D. Liu, and Z. Xiong, "Cnn-based text image super-resolution tailored for ocr," in *Proceedings of the IEEE Visual Communications and Image Processing*, 2017, pp. 1–4.

[6] R. Wongso, F. A. Luwinda *et al.*, "Evaluation of deep super resolution methods for textual images," *Procedia Computer Science*, vol. 135, pp. 331–337, 2018.

[7] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proceedings of the European Conference on Computer Vision*, 2014, pp. 184–199.

[8] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 624–632.

[9] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1664–1673.

[10] J. J. Weinman, Z. Butler, D. Knoll, and J. Feild, "Toward integrated scene text reading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 375–387, 2013.

[11] A. Bissacco, M. Cummins, Y. Netzer, and H. Neven, "Photoocr: Reading text in uncontrolled conditions," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 785–792.

[12] X. Chen, T. Wang, Y. Zhu, L. Jin, and C. Luo, "Adaptive embedding gate for attention-based scene text recognition," *Neurocomputing*, vol. 90, pp. 261–271, 2019.

[13] B. Shi, M. Yang, X. Wang, P. Lyu, C. Yao, and X. Bai, "Aster: An attentional scene text recognizer with flexible rectification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 9, pp. 2035–2048, 2018.

[14] C. Luo, L. Jin, and Z. Sun, "Moran: A multi-object rectified attention network for scene text recognition," *Pattern Recognition*, vol. 90, pp. 109–118, 2019.

[15] T. Wang, D. J. Wu, A. Coates, and A. Y. Ng, "End-to-end text recognition with convolutional neural networks," in *Proceedings of the 21st International Conference on Pattern Recognition*, 2012, pp. 3304–3308.

[16] L. Christian, T. Lucas, H. Ferenc, C. Jose, C. Andrew, A. Alejandro, A. Andrew, T. Alykhan, T. Johannes, W. Zehan *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4681–4690.

[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 2, pp. 295–307, 2015.

[18] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[19] A. Graves, S. Fernández, F. Gomez, and J. Schmidhuber, "Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks," in *Proceedings of the 23rd International Conference on Machine Learning*, 2006, pp. 369–376.

[20] R. Li, "Nonlinenr total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, pp. 259–265, 1992.

[21] J. Max, S. Karen, V. Andrea, and Z. Andrew, "Synthetic data and artificial neural networks for natural scene text recognition," *arXiv preprint arXiv:1406.2227*, 2014.

[22] M. Anand, A. Karteek, and J. CV, "Scene text recognition using higher order language priors," in *Proceedings of the British Machine Vision Conference*, 2012.

[23] K. Wang, B. Boris, and B. Serge, "End-to-end scene text recognition," in *Proceedings of the International Conference on Computer Vision*, 2011, pp. 1457–1464.

[24] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, and R. Young, "Icdar 2003 robust reading competitions," in *Proceedings of the International Conference on Document Analysis and Recognition*, 2003, pp. 682–687.

[25] K. Dimosthenis, S. Faisal, U. Seiichi, I. Masakazu, L. G. i Bigorda, S. R. Mestre, M. Joan, D. F. Mota, J. A. Almazan, and L. P. De Las Heras, "Icdar 2013 robust reading competition," in *Proceedings of the International Conference on Document Analysis and Recognition*, 2013, pp. 1484–1493.

[26] U.-V. Marti and H. Bunke, "The iam-database: An english sentence database for offline handwriting recognition," *International Journal on Document Analysis and Recognition*, vol. 5, pp. 39–46, 2002.