SINGLE IMAGE SUPER RESOLUTION USING MID-SCALED SRGAN

Harish Iyer Ganesan, Parveen Kaur, Poojitha Baskaran

Electrical and Computer Engineering, University of Waterloo, Ontario

ABSTRACT

License Plate Recognition (LPR) is an essential application of image processing and deep learning and has been focus of research for the past few decades. However, because of the low quality of images as a result of high car speeds and poor resolution of analogue cameras, many license plates may not be recognized precisely by LPR models. In order to address this problem, we have proposed an SRGAN [1] based improvised LPR model using a Mid-Scaled Super Resolution Generative Adversarial Network (MS-SRGAN) for effective super-resolution of images containing license plate. Our paper introduces architectural changes and additional loss components to the current networks for image enhancement and super resolution. The results obtained show that using MS-SRGAN for image enhancement have comparable performance with other super-resolution models, and increases the accuracy of LPR system from 26.6% to 74.5% for lower quality images.

Index Terms— License Plate Recognition, Image Enhancement, Image Super Resolution.

1. INTRODUCTION

License Plate Recognition (LPR) is a widely implemented computer vision application [2], [3], [4] which is a tool that recognizes and translates characters in a license plate from image into editable text. LPR has many practical uses such as controlling and monitoring traffic, managing parking, and security. Since the accuracy of Optical Character Recognition (OCR) is proportional to the input image quality, LPR faces several obstacles. Capturing a clear image of the License Plate is a challenge because of the car's constant motion and high speed. This is further complicated by the use of low-cost analogue cameras that provide low resolution and poor quality in most security systems. Modern LPR techniques primarily depend on only one captured image to recover the characters. These techniques, however, may not function well with images that have low resolution and quality [5], [6], [7], [8]. Studies show that image quality and resolution can be improved using CNN and deep learning algorithms [9]. Hence, using a Single Image Super-Resolution (SISR) model for image enhancement using CNN will filter and improve the image, resulting in better accuracy for LPR [10]. SISR converts a low-resolution (LR) image to a high-resolution (HR) image. The paper proposes a Mid-Scaled Super Resolution Generative Adversarial Network (MS-SRGAN) for Image Enhancement based on SRGAN [1] architecture to improve the low resolution image by implementing a deep learning system.

2. BACKGROUND AND RELATED WORK

2.1. Single Image Super Resolution

SISR has been researched and implemented for several applications in the past few decades. A deep convolutional network has lately gained much traction for its ability to map LR images to HR images [1], [11]. The first study, called SRCNN [9] aimed to lower the Mean Square Error (MSE) [12] between resolved image and ground truth by training a CNN network comprised of three convolution layers. The use of SRCNN for end-to-end mapping is concerned state of the art [12]. Multiple architectures for SISR have since been proposed.

The GAN [13] networks have been implemented widely for numerous applications, and by using an adversarial loss function, they have enabled unsupervised learning in the absence of labels. The GAN architecture is made up of two networks, the first of which is the generator, that generates the image. The discriminator is the second network, which attempts to differentiate between the generated and real images. The generator aims to mislead the discriminator by creating nearly indistinguishable visuals from the real images. SRGAN has demonstrated that adding the adversarial loss improves the quality of HR images and makes them more realistic. Instead of applying MSE directly to the rebuilt picture, the SRGAN includes the perceptual loss [14], which tries to lower the MSE between the reconstructed HR image feature mappings and the VGG [15] produced ground truth. The addition of perceptual loss produces more visually appealing images with high-frequency features than MSE-based approaches. Further, Enhanced SRGAN (ESRGAN) [11] introduces Residual-in-Residual Dense Block (RRDB) without batch normalization for network architecture and an improvement of perceptual loss utilising the features before activation, that may provide more substantial supervision for consistency in brightness, and texture recovery.

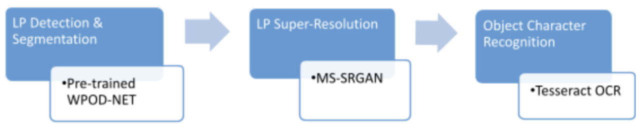


Fig. 1. Overview of the proposed LPR system

2.2. License Plate Recognition using Image Super Resolution

Many studies on licence plate recognition and approaches to increase its accuracy and efficacy have been done. Despite the fact that LPR accuracy is proportional to the quality of the image, many researchers continue to use older methods such as image processing and interpolation techniques [16], [17]. Few others concentrate on generating a HR image from several low-resolution images [18], [19]. For LR, a character semantic-based super-resolution is discussed in [20]. However, if the image is distorted with noise, segmenting the characters may become difficult. A LP SISR was proposed in [21] using SRGAN and a perceptual loss. Despite the fact that perceptual OCR loss resulted in higher character accuracy, the quality of image was no better than SRGAN. Bicubic interpolation with Gaussian noise was used to downsample the LR images from HR images, Although the low-resolution image is corrupted with this noise, it is different from the real-world images, involving very random noise.

3. SYSTEM DESIGN AND METHODOLOGY

The LPR model proposed based on a few modifications to the Super-Resolution GAN architecture (explained in 3.1) is shown in Fig.1. Using the Warped Planar Object Detection Network (WPOD-NET) [22], we derive and extract the Region of Interest (ROI) of input images from license plates dataset and then use bicubic interpolation for downsampling. The Low-Resolution (downsampled) images are fed to the generator network for training. The generator produces output in the form of super-resolved images which are subsequently input into a discriminator differentiating between the original and generated photos. Finally, the super-resolved output images are supplied to the OCR framework once the network has been trained, and Tesseract OCR [23] is used for identification of license plate characters.

3.1. Model architecture and proposed improvements

3.1.1. *Scaled image dimensions in generator*

We propose a novel architectural change to the SRGAN generator with the intuition to train the model on scaled image dimensions. Previous architecture of SRGAN focus on learning the parameters of the model on the same dimensions as of the input image, and the scaling is done at the end of the residual blocks. We modify the architecture to split the process of image scaling into three parts. As described in Fig.2, the list of B (10 in our case) residual blocks is split into two blocks (B/2), and we scale the image size midway between the two blocks. This enables the first residual block to learn the features and parameters of the image of original size and the second residual block from the scaled image.

Generator: The architecture is a combination of two residual block lists and three pixel shufflers. The model learns on both input image dimensions as well as the scaled size.

Residual Block List1: Inspired from SRGAN, the residual blocks made of a list of Convolution layers combined with Batch Normalization and ReLU activation are used. The skip connections are added among the residual blocks. This residual block list trains and learns on the input image dimensions(192x96 in our case). The combination of all the filters will give good representation of the original image.

Pixel Shuffler: The output from Residual Block List1 is fed to a convolution layer which scales the depth (number of channels) of the image 4 times. The subsequent pixel shuffler reuses the increased depth to scale the size of the image. The conversion of depth to the spatial dimensions is given by [24] as: $(Ch \times r^2, W, H)$ to $(Ch, r \times W, r \times H)$, where Ch is number of channels, and H, W refers to image's height and width respectively. This converts the quadrupled depth (2^2) to doubled image dimensions $(2 \times W, 2 \times H)$, leaving the depth of image the same as input again.

Residual Block List2: The configuration of this residual Block List is the same as the previous one with a major difference of image size on which these resnet blocks are learning which is twice the size of the original image. This is crucial for our architecture performance as we propose that the scaled dimensions of the image will infuse the learning of the features of the HR image. This enables the resnet blocks to capture complex spatial features on scaled dimensions of the HR image resulting in better representation of complex intrinsic features. Further, convolution layers and pixel shuffler are added, scaling the size of the image in two steps. The Skip connections are added through the network as shown in Fig.2 to pass on the feature captured by the previous layers directly to the further layers. It helps to prevent vanishing gradients in the deep architecture.The discriminator architecture is similar to [1].

3.1.2. *Contrastive loss function*

Inspired from [1], the generator loss is formed as the weighted sum of pixel-wise MSE Loss, Entropy loss and content loss component. We introduce a contrastive loss component to this generator loss to enhance the overall generator performance. This loss is information theory-based for determining

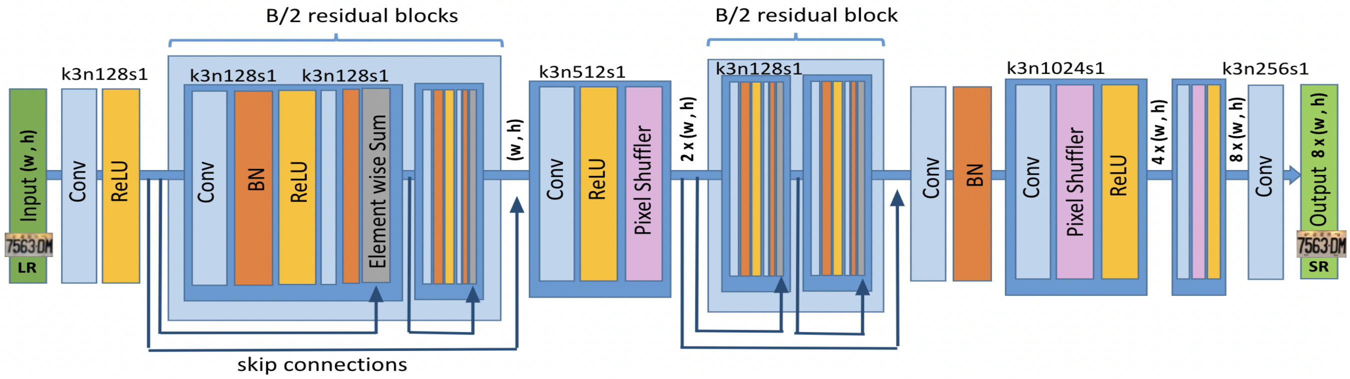


Fig. 2. Generator architecture of proposed MS-SRGAN, here k - Kernel size, Conv - Convolution Layer, n - number of filters, and s - stride.

the similarity between the original(LR) and the output of the generator (HR image). A reconstruction that has high mutual information with the ground truth is considered optimal. We use the generator output as anchor and ground truth as positive sample and input LR images as negative samples. This loss has two opposite operators: one attracts the anchor towards the positive sample while the other repels anchor away from negative samples. From [25], the contrastive loss equation is

$$L_{con}(a, a^+, a^-) =$$

$$\log \frac{exp(s(a, a^+/\tau))}{exp(s(a, a^+/\tau)) + \sum_{n=1}^{N} exp(s(a, a^-/\tau))}$$

where a^+ is the ground truth, a^- is the LR images and s ($a1$, $a2$) = $a1^T a2$ gives the cosine similarity between two encoded patch signals. If a patch is exactly reconstructed by the generator, then The numerator will be maximised because the embeddings will have perfect similarity. Thus, the updated generator loss is given as:

$$L_{genLoss} = L_{mse} + 0.2 * L_{entropy} + 0.1 * L_{content} + 0.1 * L_{con}$$

where $L_{entropy}$ is generator adversarial loss. $L_{content}$ is the perceptual loss, defined using activation layers(ReLU) of the SRGAN. The loss function focuses on perceptual similarity. The Discriminator loss is inspired from [26].

3.2. Dataset

We train and test the proposed model on the publicly available application-oriented license plate (AOLP) benchmark dataset [27] consisting 2,049 images from different locations, times, traffic conditions, and weather. Depending on a variety of application characteristics, the entire dataset is divided into three classes: Road Patrol(RP) has 611 images, Access Control(AC) has 681 images , the Traffic Low Enforcement(LE) has 757 images. Our training set consists of AC and LE images and testing set consists of RP images.

3.3. Image preprocessing and License Plate Segmentation

For detection and segmention of the license plate from images of the dataset, we use a pre-trained WPOD-NET (reusing the model implemented in [22]). The images are normalized and passed to the WPOD-NET, where it obtains license plate boundary coordinates.With the help of the gathered coordinates, a bounding box is built around the licence plate, which is then utilised to extract the licence plates from the images. Further, the segmented LP is uniformly resized to 192x96 image.

3.4. Hyper parameters and Training

The model is trained and tested using the pairs of original and downscaled images. Every pair contains a colored LR image(bicubic downsampled by 8 times) and HR image(Original Image). LR image has resolution of 24×12 while the HR image resolution is scaled eight times, which is 192×96. The Generator of the model is pretrained inspired from [28] for 200 epochs with Adam Optimizer with a learning rate of 0.001. The training of Generator and Discriminator (Min-Max game) is done for 100 epochs with a Learning rate of 0.0001. Different learning rates as 0.005,0.001, 0.0001 were analyzed and 0.0001 was finalized. The open source implementation by [29] was modified and used for training and hyperparameter tuning.

4. RESULTS AND DISCUSSION

4.1. Enhancement and Super Resolution results

We compare the performance of SRGAN [1], ESRGAN [11] and MS-SRGAN. Table 1 summarises the quantitative results of the models. The results reveal that MS-SRGAN network for the AOLP dataset achieves results that are comparable to the present SISR techniques. From Fig 4, we analyse that while HR images generated from SRGAN and ESRGAN are visually comparable, ESRGAN image also includes color information stored in original image.

Table 1. Comparison of MS-SRGAN over PSNR, and SSIM with SRGAN, and ESRGAN on AOLP dataset

	SRGAN	**ESRGAN**	**MS-SRGAN**
PSNR	23.2	22.9	21.2
SSIM	0.42	0.21	0.45

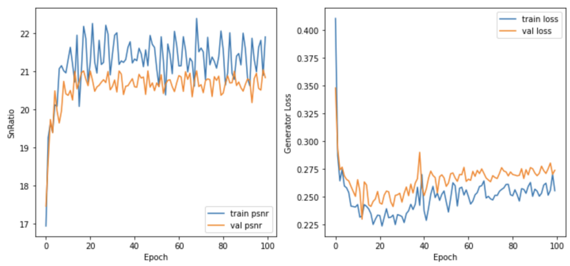


Fig. 3. PSNR and Generator loss over epochs for MS-SRGAN

For our model MS-SRGAN, performance metrics for training and validation data over epochs is shown in Fig 3, that confirms that PSNR improves over time and Generator Loss initially decreases and then slightly increases. The increase in generator loss shows the need for extensive hyperparameter tuning which could not be performed due to computational limitations.

4.2. License Plate Recognition results

The LPR results using Tesseract OCR (implemented as in [30]) before and after the use of MS-SRGAN on 100 LR images from AOLP dataset are shown in Table 2. The average accuracy (calculated as number of correct characters detected by LP total character length) rate when LR AOLP images are used has improved after using MS-SRGAN from 26.6 % to 74.5% when the Tesseract engine is used. Table 3 shows that LPR error has improved from 5 to 1 character for the same LR image before and after MS-SRGAN.

4.3. Computational comparisons

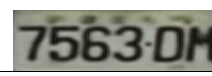
Table 3 shows a comparison of MS-SRGAN with existing SISR approaches with respect to number of trainable parameters and per epoch training and testing time duration. MS-

Fig. 4. Comparsion SRGAN,ESRGAN and MS-SRGAN

Table 2. Average LPR accuracy per image (for 100 LR AOLP images) before and after the use of MS-SRGAN

	Tesseract before SR	**Tesseract after SR**
LR AOLP	26.6%	74.5%

Table 3. LPR result comparison before and after the use of MS-SRGAN

	Input Image	**Tesseract output**	**No. of Errors**
Before MS-SRGAN		130301	5
After MS-SRGAN		753608	1

SRGAN has a lower total number of parameters than ESRGAN but more than SRGAN. The amount of parameters has a direct relationship with the number of iterations required during testing. As a result, MS-SRGAN takes longer to test than SRGAN but less than ESRGAN. The quantity and complexity of loss functions affect the processing time during training. Because of the many loss functions utilised, MS-SRGAN takes longer to train than the other techniques, making it computationally more expensive.

Table 4. Comparison of training and testing time between SRGAN, ESRGAN and MS-SRGAN

	SRGAN	**ESRGAN**	**MS-SRGAN**
Training time per step	0.6450s	0.1020s	0.8810s
Testing time per test image	0.0230s	0.0450s	0.0210s
Trainable parameters	4,438,339	6,167,747	5,471,043

5. CONCLUSION AND FUTURE WORK

In the research project, we used the Super Resolution GAN architecture (SRGAN) as the foundation for this research and proposed two enhancements: mid-scaling of the input picture in the generator and the addition of a contrastive loss function to enhance the quality of the generated SR image for License Plate image enhancement application. The results show comparable image enhancements with the existing SISR models. Future work might focus on training the model from the ground up by up-scaling predictions while computing loss, as well as incorporating the proposed modifications into existing SotA SISR models. Further, an improvised down-sampling technique can be explored to model real-life noise.

6. REFERENCES

[1] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016.

[2] Jiafan Zhuang, Saihui Hou, Zilei Wang, and Zheng-Jun Zha, "Towards human-level license plate recognition," September 2018.

[3] Nuzulha Khilwani Ibrahim, Emaliana Kasmuri, Norazira A. Jalil, Mohd Adili Norasikin, Sazilah Salam, and Mohamad Riduwan Md Nawawi, "License plate recognition (LPR): A review with experiments for malaysia case study," *CoRR*, vol. abs/1401.5559, 2014.

[4] Cheng-Hung Lin and Y. Li, "A license plate recognition system for severe tilt angles using mask r-cnn," *2019 International Conference on Advanced Mechatronic Systems (ICAMechS)*, pp. 229–234, 2019.

[5] Sérgio Montazzolli and Claudio Jung, "License plate detection and recognition in unconstrained scenarios," 09 2018.

[6] Zhongze Zhang and Yi Wan, "Improving the accuracy of license plate detection and recognition in general unconstrained scenarios," *2019 IEEE Symposium Series on Computational Intelligence (SSCI)*, pp. 1194–1199, 2019.

[7] Naiguo Wang, Xiangwei Zhu, and Jian Zhang, "License plate segmentation and recognition of chinese vehicle based on bpnn," *2016 12th International Conference on Computational Intelligence and Security (CIS)*, pp. 403–406, 2016.

[8] Aiswarya Menon and Bini Omman, "Detection and recognition of multiple license plate from still images," 12 2018, pp. 1–5.

[9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295–307, feb 2016.

[10] Younkwan Lee, Jiwon Jun, Yoojin Hong, and Moongu Jeon, "Practical license plate recognition in unconstrained surveillance systems with adversarial super-resolution," 01 2019, pp. 68–76.

[11] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang, "ESRGAN: enhanced super-resolution generative adversarial networks," *CoRR*, vol. abs/1809.00219, 2018.

[12] Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz, "Loss functions for image restoration with neural networks," *IEEE Transactions on Computational Imaging*, vol. 3, no. 1, pp. 47–57, 2017.

[13] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial networks," 2014.

[14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," *CoRR*, vol. abs/1603.08155, 2016.

[15] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2015.

[16] Menna Ghoneim, Mohamed M. Rehan, and Hisham Othman, "Using super resolution to enhance license plates recognition accuracy," *2017 12th International Conference on Computer Engineering and Systems (ICCES)*, pp. 515–518, 2017.

[17] Jie Yuan, Sidan Du, and Xiang Zhu, "Fast super-resolution for license plate image reconstruction," *2008 19th International Conference on Pattern Recognition*, pp. 1–4, 2008.

[18] Vojtech Vasek, Vojtech Franc, and Martin Urban, "License plate recognition and super-resolution from low-resolution videos by convolutional neural networks," in *BMVC*, 2018.

[19] Gabriele Guarnieri, Marco Fontani, Francesco Guzzi, Sergio Carrato, and Martino Jerian, "Perspective registration and multi-frame super-resolution of license plates in surveillance videos," *Forensic Science International: Digital Investigation*, vol. 36, pp. 301087, 03 2021.

[20] Yuexian Zou, Yi Wang, Wenjie Guan, and Wenwu Wang, "Semantic super-resolution for extremely low-resolution vehicle license plate," *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3772–3776, 2019.

[21] Seyun Lee, Ji-Hwan Kim, and Jae-Pil Heo, "Super-resolution of license plate images via character-based perceptual loss," *2020 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 560–563, 2020.

[22] Sameer Bairwa, "automatic license plate number detection," https://github.com/sameerbairwa/ANPR?ref=hackernoon.com, 2020.

[23] R. Smith, "An overview of the tesseract ocr engine," in *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, 2007, vol. 2, pp. 629–633.

[24] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan 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.

[25] Alex Andonian, Taesung Park, Bryan Russell, Phillip Isola, Jun-Yan Zhu, and Richard Zhang, "Contrastive feature loss for image prediction," 11 2021.

[26] Taizhang Shang, Qiuju Dai, Shengchen Zhu, Tong Yang, and Yandong Guo, "Perceptual extreme super resolution network with receptive field block," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1778–1787, 2020.

[27] Gee-Sern Hsu, Jiun-Chang Chen, and Yu-Zu Chung, "Application-oriented license plate recognition," *IEEE transactions on vehicular technology*, 2012.

[28] Ian Goodfellow, "Nips 2016 tutorial: Generative adversarial networks," *arXiv preprint arXiv:1701.00160*, 2016.

[29] Zixuan Zhang and Chengxuan Cai, "License plate enhancement - from tv shows to reality," https://github.com/zzxvictor/License-super-resolution, 2020.

[30] Smeet Kevadiya, "License plate recognition with opencv and tesseract ocr," https://www.geeksforgeeks.org/license-plate-recognition-with-opencv-and-tesseract-ocr/, 2020.