CS766 Project Proposal

Zhikang Hao
UW-Madison
hao45@wisc.edu

Wendi Li
UW-Madison
wendi.li@wisc.edu

Zhikang Meng
UW-Madison
zmeng34@wisc.edu

1 INTRODUCTION

Super-resolution tasks are widely used in generating high-quality images or videos from previous low-resolution sources to fit current high-resolution screens, removing blurs, or improving the image classification dataset. Photos/videos distributed decades ago usually have very low resolution due to the hardware restriction on-screen and storage. Even now, the camera on the phone is not powerful enough to reach high resolution with optical zooming. Thus, providing higher resolution without distortion is needed.

Single Image Super-Resolution (SISR) [11] is one of the most difficult super-resolution tasks that target using a single degraded/low-resolution image to generate a clear high-resolution image. The common degradation that makes the image unclear includes blur, noise, and compression. Traditional method without machine learning works poorly on this task.

Until, not surprisingly, Deep Learning based methods such as CNN and GAN provide a relatively promising solution to the SISR problem and achieve high scores on such tasks [1]. In 2014 SRCNN[2] was proposed for the SISR task and perform better than previous non-DL algorithms. Currently, SRGAN and its variation are mainstream in generating high-resolution images such as high-res face generation in either images or videos[4].

However, the deep learning-based implementation on the noisy data usually faces generalization difficulties that trained models cannot be well applied to images from the real world. The largest difficulty lies in the fact that real-world noise models are unpredictable. Several works such as ESRGAN[7] aim for solving this issue but it is still an open problem. Today various camera and image transmissions may cause large variance on the image noise model, which is further impacted by various image/video compression. If one takes a photo and uploads it to social media, the photo can be compressed two times on both ends. Even with the blind approach[5] whose model does not depend on prior knowledge of image type, although provides improvements, still makes certain assumptions on the degradation. And lacking perfectly paired HR/SR data further influences the performance of deep learning-based methods.

2 RELATED WORKS

RSR [1] raises another approach to solving the generalization issue by employing adversarial attacks on SR models, generating outliers of noise distributing, and training for robustness. The author tested RSR on real-world datasets FACE [10] and DPED [3]. The result shows this approach achieves better results than ESRGAN.

Real-ESRGAN [6] introduces a more complex degradation model targeting real-world image degradation within networks, where the images or videos are compressed or resized multiple times. Compared with ESRGAN, it performs well on online images or videos but still face generalization issue when shifting to other domain such as street view.

TTSR [9] uses the transformer model to bring attention mechanism to the SISR task. This approach adds image texture transfer and synthesis functions into the pipeline after feature extraction by DNN. On the generalization test, the authors show transformer approach outperforms a series of GANs in generalization tests with a better understanding of contextual texture.

3 PROBLEM SETUP

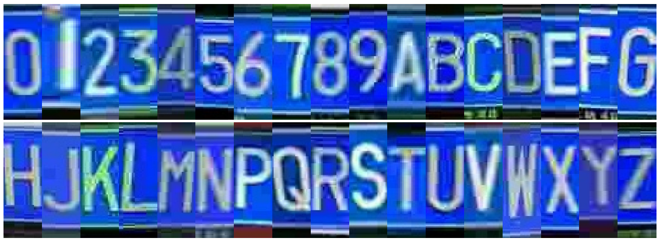
One possible SISR task is Vehicle license plate recognition. The training environment of a standard plate recognition classifier could be substantially different from the real-world environment. Characters in training data sets are often clear and easy to identify. Models build on these are hard to achieve strong generalization capability. License plates pictures that are captured on dark rainy days or snowing days are usually fuzzy, blurred, and full of noise. These low-resolution pictures really hurt the accuracy of standard recognition classifiers. After applying ESRGAN to generate high-resolution pictures from these low-resolution pictures, we can make characters clear and more recognizable. Plate classifiers use these high-resolution pictures to make predictions and improve overall accuracy. The problem we want to solve here is how to build our SRGAN model to generate high-resolution plate pictures from real-world low-quality plate pictures so that the classifier can reach its best accuracy.

4 METHODOLOGY

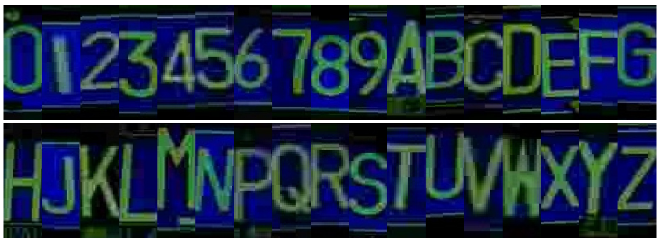
The Chinese City Parking Dataset (CCPD) [8] that contains 250 thousand license plate pictures will be used. The CCPD images are sampled from different environments, the description of this CCPD dataset is described in Table 1. These various environments are suitable for us to evaluate whether the SISR methods works well on such scenarios.

The overall process of this experiment is as follows:

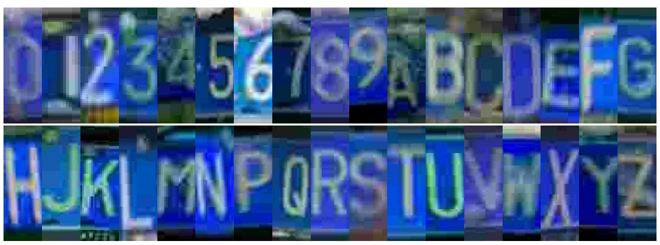
(1) Build a plate classifier based on traditional SVM-based methods and use pictures from CCPD-Base.
(2) Build ESRGAN high-resolution picture generating model using pictures from low-resolution CCPD-DB, CCPD-Weather, CCPD-Fn, and CCPD-Challenge.
(3) Feed above resolution pictures into the plate classifier from 1 and evaluate its accuracy. The accuracy will be measured as the correct prediction percentage.
(4) Iteratively improve the model we made in (2) and evaluate its accuracy in (3) until reaching the best accuracy.



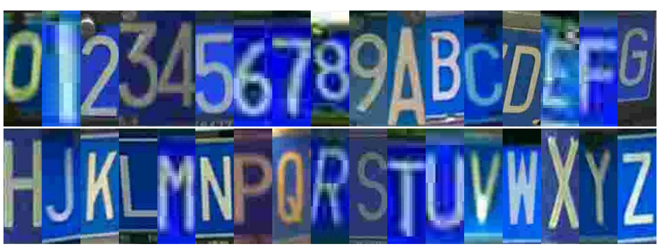
(a) License plate samples in *CCPD-Base*.



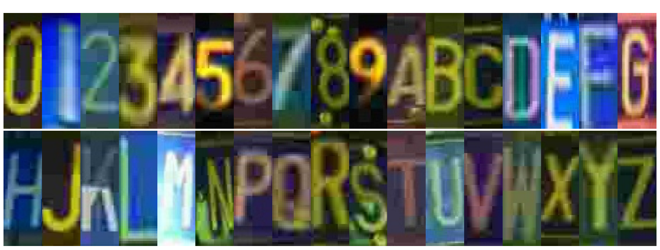
(b) License plate samples in *CCPD-DB*.



(c) License plate samples in *CCPD-Weather*.



(d) License plate samples in *CCPD-Fn*.



(e) License plate samples in *CCPD-Challenge*.

Figure 1: Five types of license plates compared.

Table 1: THE DESCRIPTION OF DATA SET

Type	Description
CCPD-Base	The only common feature of these photos is the inclusion of a license plate.
CCPD-DB	Illuminations on the LP area are dark, uneven or extremely bright.
CCPD-Weather	Images taken on a rainy, snow, or fog day.
CCPD-Fn	The license plate is relatively far or near to the camera
CCPD-Challenge	The most challenging images

The pictures in *CCPD-Base* are the collection of most common ones. They are shot right in front of the car when the light is relatively abundant. The angle of the shooting is relatively positive, the focus is good, and there is almost no blur.

5 TIMELINE

Feb 28	finalize reference resources
March 12	build standard plate classifier
March 30	build best high-resolution picture generator
April 10	finalize implementation and evaluate overall performance
April 24	finalize report and presentation slides

REFERENCES

[1] Angela Castillo, María Escobar, Juan C. Pérez, Andrés Romero, Radu Timofte, Luc Van Gool, and Pablo Arbelaez. 2021. Generalized Real-World Super-Resolution Through Adversarial Robustness. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*. 1855–1865.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. 2014. Learning a Deep Convolutional Network for Image Super-Resolution. In *Computer Vision – ECCV 2014*, David Fleet, Tomas Pajdla, Bernt Schiele, and Tinne Tuytelaars (Eds.). Springer International Publishing, Cham, 184–199.

[3] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. DSLR-Quality Photos on Mobile Devices with Deep Convolutional Networks. arXiv:1704.02470 [cs.CV]

[4] Junjun Jiang, Chenyang Wang, Xianming Liu, and Jiayi Ma. 2021. Deep Learning-based Face Super-Resolution: A Survey. arXiv:2101.03749 [cs.CV]

[5] Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. 2021. Blind Image Super-Resolution: A Survey and Beyond. arXiv:2107.03055 [cs.CV]

[6] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. 2021. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data. arXiv:2107.10833 [eess.IV]

[7] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Chen Change Loy, Yu Qiao, and Xiaoou Tang. 2018. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. arXiv:1809.00219 [cs.CV]

[8] Zhenbo Xu, Wei Yang, Ajin Meng, Nanxue Lu, and Huan Huang. 2018. Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline. In *Proceedings of the European Conference on Computer Vision*. 255–271.

[9] Fuzhi Yang, Huan Yang, Jianlong Fu, Hongtao Lu, and Baining Guo. 2020. Learning Texture Transformer Network for Image Super-Resolution. arXiv:2006.04139 [cs.CV]

[10] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. 2016. WIDER FACE: A Face Detection Benchmark. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 5525–5533. https://doi.org/10.1109/CVPR.2016.596

[11] Wenming Yang, Xuechen Zhang, Yapeng Tian, Wei Wang, Jing-Hao Xue, and Qingmin Liao. 2019. Deep Learning for Single Image Super-Resolution: A Brief Review. *IEEE Transactions on Multimedia* 21, 12 (Dec 2019), 3106–3121. https://doi.org/10.1109/tmm.2019.2919431