LP-Diff: Towards Improved Restoration of Real-World Degraded License Plate

Supplementary Material

A. Dataset

In this section, we provide additional information on the MDLP, building on the main script. The MDLP contains a total of 10,245 data groups. Each group contains several consecutive frames, ranging from degraded to clear images. Figure 4 displays an example group, where we select every 10th frame for illustration due to the large number of frames.

We provide some statistics to better illustrate MDLP. The first character of the license plate represents the province of the vehicle, so we can categorize the MDLP by the province. The statistical results are shown in Figure 1. Although some categories have a relatively low proportion, the different categories represent only the variation in the first character and do not impact the model's learning of letters and digits. For categories with a larger proportion, the model can learn the features of the first character more stably, while for categories with fewer samples, it serves as a good test for the model's few-shot learning capability. Additionally, the images in the MDLP are captured under different lighting conditions, which we categorized as low-light and normal-light environments, as shown in Figure 2.

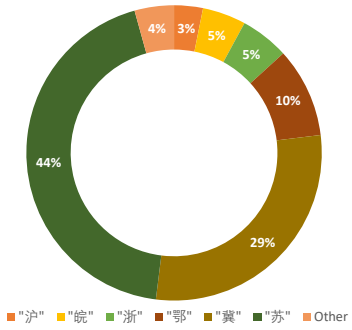

Figure 1. Composition of the license plate dataset MDLP. Categorized by the province of the license plate.

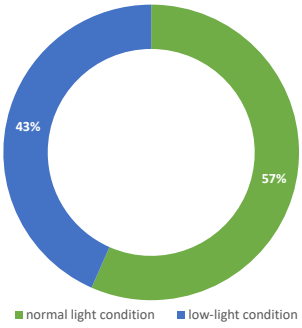

Figure 2. Composition of the license plate dataset MDLP. Categorized by the lighting conditions.

In addition to image data, we also provide the textual content of license plates to support future work and research involving multi-modal information. We use a pre-trained CRNN [4] model to recognize the license plates in the clear images and then manually verify the results to ensure the accuracy of the content.

For ethical and legal considerations, we remove all privacy-related information from the images in MDLP, including faces and surrounding scenery, leaving only the license plate numbers. Additionally, we also remove sensitive metadata, such as GPS location and timestamps.

Figure 3 shows the pipeline of dataset collection and post-processing.

B. More Results

In this section, we provide additional comparison results. As shown in Figure 5, we present enlarged input images of three consecutive degraded frames: f_1, f_2, and f_3, with the recognition results from the license plate recognition model displayed above the images. Incorrectly recognized characters are marked in red. It can be observed that most other SOTA models exhibit more than three-character errors, while LP-Diff demonstrates higher restoration quality for individual characters compared to the others.

Upon analyzing the experimental results, we find that our LP-Diff more accurately restores certain specific characters that are commonly misrecognized in real-world license plate recognition models. These challenging characters include 'Q', 'D', 'U', '2', 'W', 'T', 'V', and 'B'. Due to their structural similarities with other characters, these letters often pose challenges for other SOTA models in license plate restoration tasks. However, with the help of the Texture Enhancement Module, LP-Diff more effectively restores these confused characters. In Figure 6, we illustrate

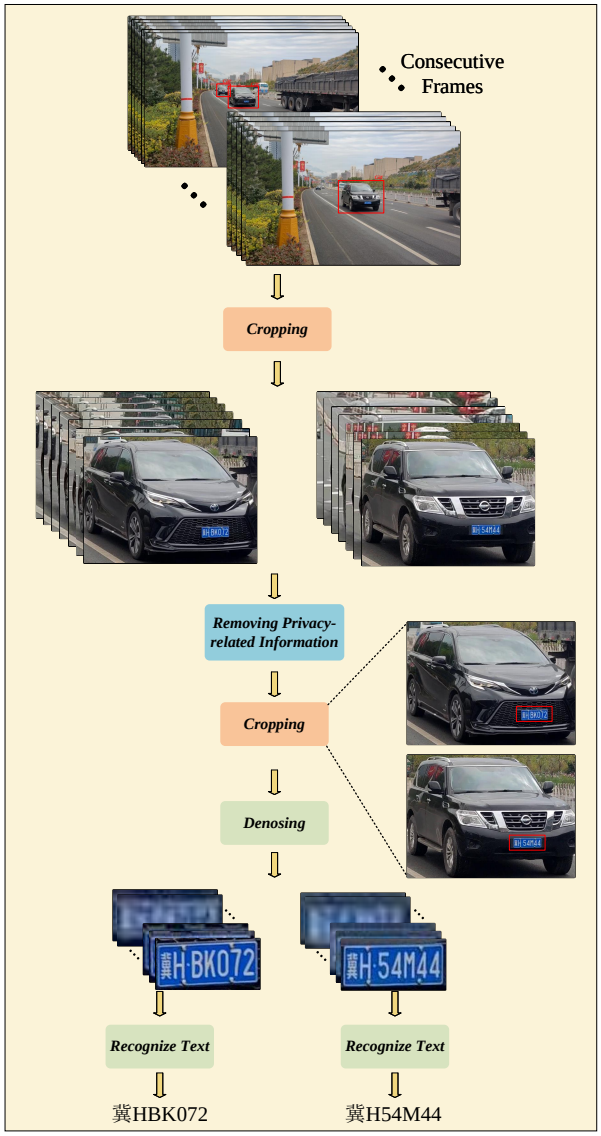

Figure 3. The pipeline of data collection and post-processing.

the results involving these characters, with the key characters highlighted in red boxes. Specifically, pairs such as <Q, 0>, <D, 0>, <U, 0>, <2, Z>, <W, X>, <T, 7>, <V, Y>, <B, 8>, exhibit structural similarities in degraded images. Figure 6 shows the confusion exhibited by other SOTA models, whereas LP-Diff delivers superior performance in correctly restoring these characters.

Moreover, LP-Diff demonstrates a stronger capability for restoring structurally complex characters lacking in the training set. Since some complex characters account for a small proportion of the training data, we specifically assess the model's ability to restore these characters to evaluate its generalization capacity. As shown in Figure 7, other SOTA models fail to accurately reconstruct the structure of these characters and mistakenly map them to similar characters that are more common in the training set. Under the same training conditions, LP-Diff is able to restore the intricate structure of these underrepresented characters with higher accuracy, demonstrating its generalization ability and few-shot learning capability.

References

[1] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22367–22377, 2023. 3

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015. 3

[3] Shuyao Shang, Zhengyang Shan, Guangxing Liu, LunQian Wang, XingHua Wang, Zekai Zhang, and Jinglin Zhang. Resdiff: Combining cnn and diffusion model for image super-resolution. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 8975–8983, 2024. 3

[4] Baoguang Shi, Xiang Bai, and Cong 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*, 39(11): 2298–2304, 2016. 1

[5] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1905–1914, 2021. 3

[6] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in Neural Information Processing Systems*, 36, 2024. 3

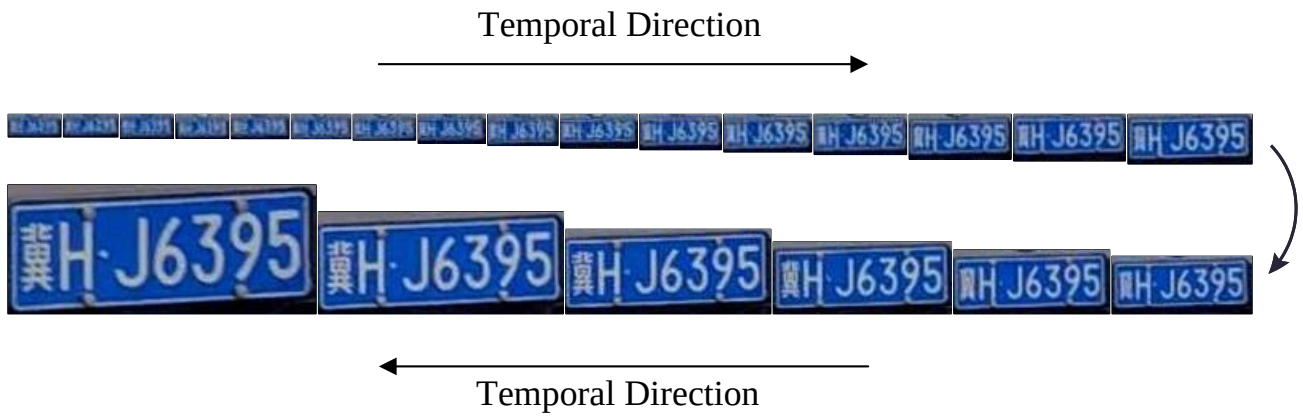

Figure 4. An example of a dataset group, which contains extensive consecutive frames, ranging from degraded to clear images. We select every 10th frame for illustration due to the large number of frames.

Figure 5. Qualitative comparison with other SOTA methods, including SRCNN [2], HAT [1], Real-ESRGAN [5], ResDiff [3], and ResShift [6] on MDLP dataset.

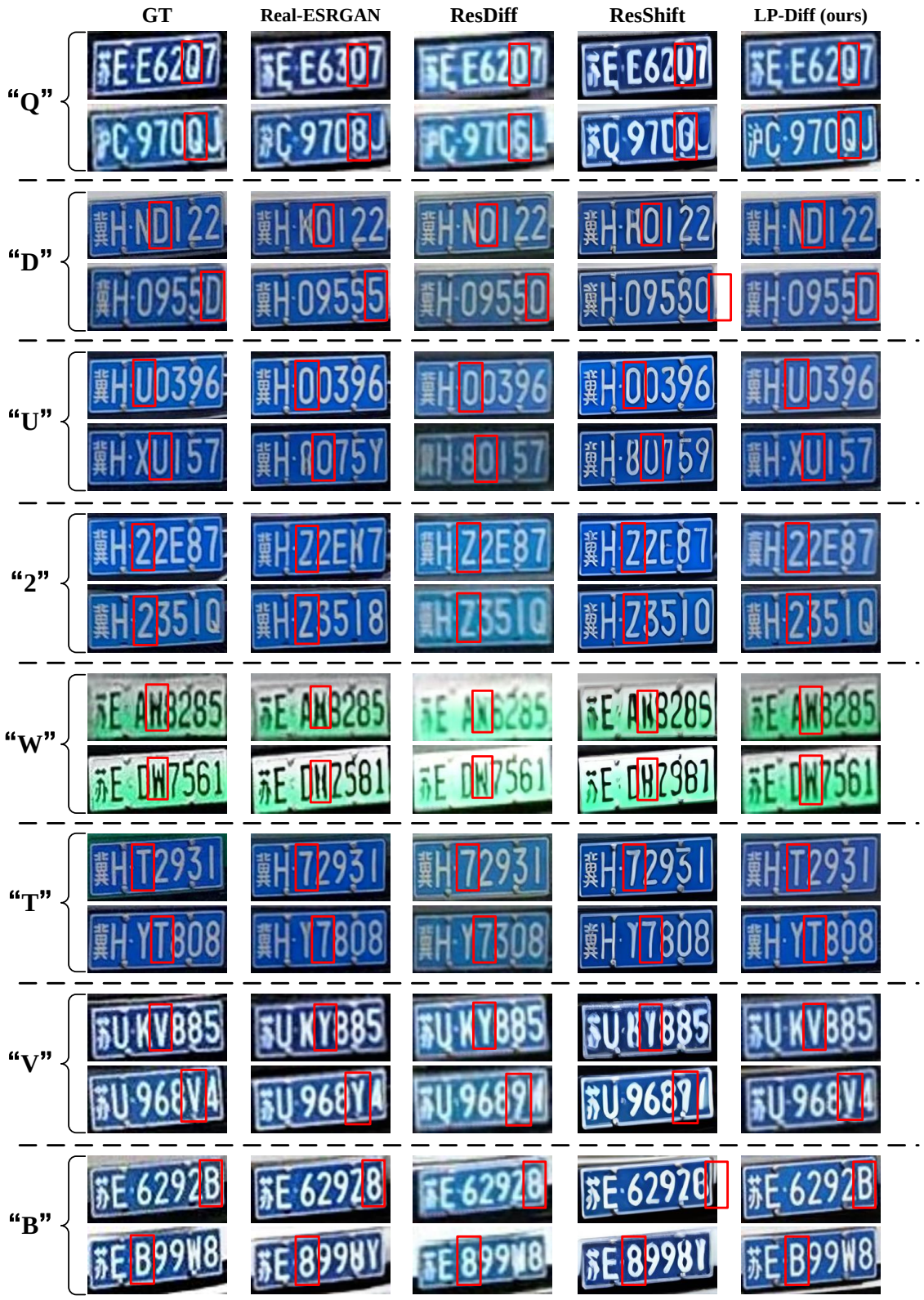

Figure 6. LP-Diff can accurately reconstruct characters that pose challenges for other SOTA models. These characters include ‘Q’, ‘D’, ‘U’, ‘2’, ‘W’, ‘T’, ‘V’, and ‘B’. The key characters are highlighted in red boxes.

Figure 7. For complex structured characters with a small number of training samples, LP-Diff can more accurately reconstruct their structure, demonstrating the generalization and few-shot learning capability of LP-Diff.