Super-Resolution Towards License Plate Recognition

Valfride Nascimento, Rayson Laroca, David Menotti
Department of Informatics, Federal University of Paraná (UFPR), Curitiba, Brazil
{vwnascimento, rblsantos, menotti}@inf.ufpr.br

***Abstract*—Recent years have seen significant developments in license plate recognition through the integration of deep learning techniques and the increasing availability of training data. Nevertheless, reconstructing license plates from low-resolution surveillance footage remains challenging. To address this issue, we propose an attention-based super-resolution approach that incorporates sub-pixel convolution layers and an Optical Character Recognition (OCR)-based loss function. We trained the proposed architecture on synthetic images created by applying heavy Gaussian noise followed by bicubic downsampling to high-resolution license plate images. Our results show that the proposed approach for reconstructing these low-resolution images substantially outperforms existing methods in both quantitative and qualitative measures. Our source code is publicly available at https://github.com/valfride/lpr-rsr-ext/.**

I. Introduction

Super-resolution is a method for enhancing the quality of an image or video by increasing its resolution. It has become a widespread technology in fields like medical imaging and surveillance [2], [3]. In recent times, there have been remarkable advancements in super-resolution techniques, particularly in interpolation-based, example-based, and deep learning-based methods [4]–[6]. These improvements have made it feasible to enhance low-resolution (LR) images and videos in a manner that was once considered impossible.

Despite advances in recent years, super-resolution remains a challenging issue due to its ill-posed nature, where there can be numerous solutions in the high-resolution (HR) space [3], [4]. Furthermore, the computational difficulty of the problem grows as the upscale factor increases, and LR images may lack sufficient information to reconstruct the desired details [3], [4]. This study focuses on the application of single-image super-resolution in the context of license plate recognition, as images from real-world surveillance systems are often characterized by low resolution and poor quality [7]–[9]. Although such challenging conditions are common in forensic applications, recent studies in license plate recognition have mainly concentrated on scenarios where the license plates are perfectly legible [10]–[14].

To address the super-resolution problem, many researchers have proposed approaches based on convolutional neural networks [3], [15], [16]. These approaches have achieved exceptional results, but often rely on deep architectures that can be computationally expensive and focus on increasing the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) without considering the particular application at hand. In the context of license plate recognition, we assert that such methods may not be effective in dealing with confusion between closely resembling characters, such as ‘Q’ and ‘O’, ‘T’ and ‘7’, ‘Z’ and ‘2’, and other similar pairs.

In this work, we present a novel approach for improving license plate super-resolution by incorporating sub-pixel convolution layers and a Pixel Level Three-Fold Attention Module (PLTFAM). Our method extends the work of Mehri et al. [16] by taking into account not only the image's pixel intensity values but also structural and textural information. To further enhance the performance, we incorporate an autoencoder that extracts shallow features by squeezing and expanding the network constructed with *PixelShuffle* (PS) and *PixelUnshuffle* (PU) layers. Additionally, we leverage a pre-trained Optical Character Recognition (OCR) model [17] to extract features from the license plate images during the training phase, resulting in improved super-resolution performance and recognition rates.

In summary, the main contributions of this work are:

- A super-resolution approach that incorporates subpixel-convolution layers in combination with a PLTFAM;
- A novel perceptual loss that combines features extracted by an OCR model [17] with L1 loss to reconstruct characters with the most relevant characteristics. This loss function allows the use of any OCR model for license plate recognition;
- The datasets we built for this work, as well as the source code, are publicly available to the research community.

The rest of this work is organized as follows. Section II provides a concise overview of relevant studies on super-resolution and license plate recognition. Section III elaborates on our proposed network architecture and the new perceptual loss function. Section IV presents the conducted experiments and results. Section V summarizes the findings, concluding this study.

II. Related Work

This section offers a concise overview of the relevant literature.

A. Single-Image Super-Resolution

Significant advancements have been made in single-image super-resolution, with the introduction of the SRCNN [18] in 2016. SRCNN outperformed early methods but had limitations such as relying on pre-upsampled LR images [19], [20]. Later

This work is a summary of a master's dissertation [1].

The second author served as an informal co-advisor for this work.

studies incorporated upsampling near the end of the network architecture to address these limitations [21], [22].

Shi et al. [22] emphasized learnable upscaling and introduced sub-pixel convolution layers for improved performance. Attention mechanisms were introduced to enhance image reconstruction, including first-order statistical attention mechanisms [23] and improved versions using second-order statistics [24]. Huang et al. [25] proposed an attention network that preserves detail fidelity by using a divide-and-conquer strategy.

In 2021, Mehri et al. [16] introduced Multi-Path Residual Network (MPRNet), which leverages information from both inner-channel and spatial features using a Two-fold Attention Module (TFAM). MPRNet has showed superior or competitive performance compared to multiple state-of-the-art methods [26]–[28]. Lastly, Zhang et al. [29] recently proposed the dual-coordinate direction perception attention mechanism for structure- and texture-preserving image super-resolution, resulting in improved image quality. As detailed in Section III, we incorporate this mechanism into the proposed approach to achieve a clearer distinction between the license plate's background and characters, thus improving font restoration.

B. Super-Resolution for License Plate Recognition

The goal of license plate recognition is to precisely extract and identify characters from each license plate. Despite recent progress and successful outcomes in license plate recognition [10], [12], [13], [30], most of the models proposed have only been trained and evaluated on HR images, where the license plate characters are clear and easily recognizable to the human eye. This does not reflect the typical conditions encountered in real-world surveillance scenarios, where images frequently have low resolution and poor quality [7]–[9].

Factors such as camera distance, motion blur, lighting conditions, image compression, and the use of cost-effective systems with rolling shutter technology can affect the quality of the license plate images [7], [31]. Super-resolution techniques have been proposed as a solution to this issue. Early approaches performed poorly on noisy images [32], [33], while recent works have shown promising results through deep learning.

Lin et al. [34] proposed a super-resolution approach for license plate recognition using Super-Resolution Generative Adversarial Networks (SRGAN) and perceptual OCR loss. Hamdi et al. [35] proposed a GAN-based architecture named Double Generative Adversarial Networks for Image Enhancement and Super Resolution (D_GAN_ESR), which outperformed previous SRGAN methods [34]. They evaluated their method using PSNR and SSIM, but acknowledged that these metrics alone may not indicate superior image reconstruction. Lee et al. [36] designed a GAN-based super-resolution model that incorporates a perceptual loss composed of intermediate features extracted by a scene text recognition model. While their method produced promising results, they did not make the dataset used available, and the degradation method employed was not detailed.

Most works evaluate quality subjectively or using metrics such as PSNR and SSIM, which have limited correlation with human assessment [37], [38]. Moreover, many previous studies performed experiments exclusively on private datasets [8], [35], [36], making accurate assessment challenging.

III. Proposed Approach

This section presents our super-resolution approach for enhancing feature extraction from low-resolution license plates. We incorporate ideas from [29] to better capture structural and textural information from the license plate images. We also introduce a novel perceptual loss function that leverages an OCR model.

A. Network Architecture Modifications

As illustrated in Fig. 1, the proposed approach for license plate super-resolution builds upon the network architecture of Mehri et al. [16] and Zhang et al. [29]. It includes a Shallow Feature Extractor (SFE), Residual Dense Blocks (RDBs), a Feature Module (FM) module, and a Reconstruction Module (RM). In a nutshell, the RM combines the output of the FM module with two long-skip connections, one from the end of the SFE module and the other from the input image, to produce the high-resolution output. Our specific modifications are discussed in the following paragraphs.

The SFE block includes a 5×5 kernel Conv. layer followed by an autoencoder that employs depthwise-separable convolutional layers (DConvs), PU, and PS operations instead of conventional Conv. layers, pooling, and upscale operations. The output of the layers is then combined with a skip connection from the initial Conv. layers and processed by the RDBs.

In Fig. 2, we present our modifications to the MPRNet's TFAM [16] to create the PLTFAM. The design of this module is based on the following insights: **(i)** images are composed of the relationship between channels, where each channel contributes unique characteristics to form the final image, therefore, the extraction of these features is crucial for proper image restoration; **(ii)** the positional information of these essential features from the channels composing the images is required; **(iii)** traditional downscale and upscale operations rely on translational invariance and interpolation techniques, which are not able to learn a custom process for different tasks; **(iv)** the module captures salient structure from the character fonts of the license plate, highlighting both structure and textural features in the image.

The Channel Unit (CA) module identifies and preserves relevant inter-channel relationship features by utilizing two parallel Conv. layers. Their outputs are concatenated and processed through convolutional, concatenating their outputs, PU, PS, and DConv layers, summarizing the inter-channel relationship features and enhancing image restoration.

The Positional Unit (POS) complements the CA module by determining the location of important features in the image. It extracts first-order statistics through pooling operations, combines the results, and processes them through DConvs and PS layers to restore the original feature map dimensions. This highlights the positions of relevant inter-channel relationship features and improves image restoration.

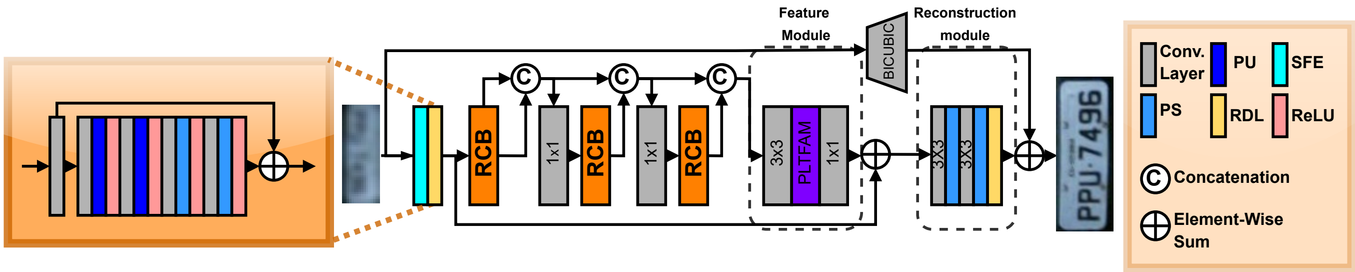
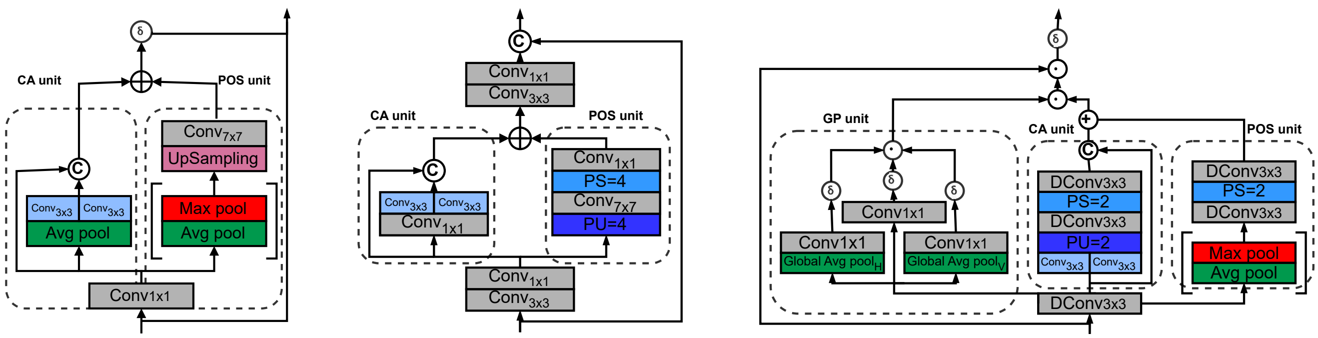


Fig. 1. The proposed architecture, which incorporates an autoencoder consisting of PS and PU layers for feature compression and expansion, respectively. This design aims to eliminate less significant features. The TFAM modules were replaced with PLTFAM modules throughout the network. PS = PixelShuffle; PU = PixelUnshuffle; RDL = Residual Dense Layers; SFE = Shallow Feature Extractor.



(a) (b) (c)

Fig. 2. Comparative illustration of the (a) Two-Fold Attention Module in MPRNet [16], (b) PixelShuffle Two-Fold Attention Module in [39] (a preliminary version of this work), and (c) PixelShuffle Three-Fold Attention Module (ours).

To enhance the network's ability to extract critical structural, textural, and geometric features from the license plate, we introduced the Geometrical Perception Unit (GP) branch. Inspired by [29], it employs global average pooling in both vertical and horizontal directions. The output undergoes point-wise convolution and sigmoid function operations, followed by element-wise multiplication to obtain the final output.

The CA, POS, and GP units' outputs are combined through element-wise sum and multiplication operations, forming the attention mask. This mask enhances the input to the PLTFAM module, effectively emphasizing key image features like inter-channel relationships, positional information, and structural details. This leads to improved image restoration.

The Residual Concatenation Blocks (RCBs) were improved by adding the PLTFAM and dilated convolution layers to the bottleneck path of the Adaptive Residual Blocks (ARB). This modification maintains the structure described in [16], while incorporating a wider context with an increased receptive field and preserving fine details in license plate images.

Returning our attention to Fig. 1, a reconstruction module was added as an output block for better aggregating fine details. It consists of two PS with a scale factor of 2, followed by DConv layers and consecutive RDBs.

B. Perceptual Loss

The proposed approach incorporates a perceptual loss function to further enhance the accuracy of the super-resolution method for license plate recognition. The perceptual loss function is designed to consider the features expected by an OCR model, which improves the system's accuracy.

$$PL = \frac{1}{n}\Big(\sum_{i=1}^{n}(H_i - S_i)^2 + \sum_{i=1}^{n}|f_{OCR}(H_i) - f_{OCR}(S_i)|\Big) \quad (1)$$

The perceptual loss function, as defined in Eq. (1), consists of two terms: the Mean Squared Error (MSE) term and the feature extraction term. The MSE term measures the pixel value difference between the HR (H_i) and super-resolution (S_i) license plate images. The feature extraction term compares the feature representations of the HR and super-resolution images obtained from an OCR model denoted as $f_{OCR}(\cdot)$.

The loss function can accommodate any OCR model for license plate recognition, providing flexibility to incorporate novel models as they become available. In this work, the multi-task model proposed by Gonçalves et al. [17] was explored due to its impressive performance and efficiency in previous research [7], [39].

The MSE term penalizes significant errors between the expected and generated images, effectively improving overall image quality and preserving important structural information. On the other hand, the feature extraction term, measured using the L1 loss, promotes robustness to noise and outliers while preserving sharp edges in the generated images. This combination of MSE and L1 loss allows for a comprehensive evaluation of the generated images, striking a balance between preserving structural information and minimizing errors.

IV. EXPERIMENTS

Here, we detail the steps taken to validate the effectiveness of our proposed method for license plate super-resolution. We first describe our experimental setup and then proceed to provide a comprehensive analysis of the results obtained.

A. Setup

We made use of license plate images obtained from the RodoSol-ALPR [30] and PKU [40] datasets. To the best of our knowledge, there is currently no public dataset that provides paired LR and HR images from real-world settings. Hence, we opted for these two datasets since they provide a wide range of scenarios under which the images were acquired.

The RodoSol-ALPR dataset consists of 20,000 images, including vehicles with Brazilian license plates and Mercosur license plates[1]. It offers a diverse range of scenarios, including variations in license plate colors, lighting conditions, and character fonts (see Fig. 3). We followed the standard protocol defined in [30], allocating 40% of the images for training, 20% for validation, and 40% for testing purposes.

Fig. 3. Some license plate images from the RodoSol-ALPR dataset [30]. The first two rows show Brazilian license plates, while the last two rows show Mercosur license plates. For scope reasons, we conduct experiments on license plates that have all characters arranged in a single row (i.e., 10K images).

The PKU dataset includes images grouped into G1-G5 representing various scenarios in mainland China, such as highways during the day (G1) and crosswalk intersections during the day or night (G5). Our experiments focused on G1-G3, totaling 2,253 images with annotated license plate text [42]. The license plate images in the PKU dataset demonstrate high quality and legibility, as shown in Fig. 4. We followed the approach of [41], [42], splitting 60% of the images for training and validation, and the remaining 40% for testing. To avoid bias, we grouped *near-duplicates* (distinct images of the same license plate) together, as recommended by Laroca et al. [43].

The HR images used in our experiments were created as follows. For each image from the datasets, we first cropped the license plate region using the annotations provided by the authors. Afterward, we used the same annotations to rectify each license plate image so that it becomes more horizontal, tightly bounded, and easier to recognize. The rectified image is the HR image.

To generate LR versions of each HR image, we simulated lower-resolution effects based on [44]. We applied iterative

[1]Following [12], [39], [41], we use the term "Brazilian" to refer to the layout used in Brazil prior to the adoption of the Mercosur layout.

Fig. 4. Examples of license plate images from the PKU dataset [40]. Although the license plates in this dataset have varying layouts, they all have seven characters.

random Gaussian noise to each HR image until reaching the desired degradation level for a LR image (i.e., SSIM < 0.1). To maintain the aspect ratio of the LR and HR images, we apply padding before resizing them to 20×40 pixels, resulting in an output shape of 80×160 pixels for an upscale factor of 4. Examples of the generated license plate images for the RodoSol-ALPR and PKU datasets are shown in Fig. 5 and Fig. 6, respectively.

Fig. 5. Some HR-LR image pairs created from the RodoSol-ALPR dataset.

Fig. 6. Examples of HR-LR image pairs created from the PKU dataset.

Our experiments were conducted using the PyTorch and Keras frameworks on a high-performance computer featuring an AMD Ryzen 9 5950X CPU, 128 GB of RAM, and an NVIDIA Quadro RTX 8000 GPU with 48 GB of memory.

We utilized the Adam optimizer with a learning rate of 10^{-4}. The learning rate decreased by a factor of 0.3 (down to 10^{-7}) when no improvement in the loss function was observed. The training process was terminated after 20 epochs without any decrease in the loss function.

B. Experimental Results

In the license plate recognition literature, models are typically evaluated by the ratio of correctly recognized license plates to the total number of license plates in the test set [12], [13], [43]. A license plate is considered correctly recognized if all characters are identified accurately. Considering our focus on low-resolution license plates, which are very common in forensic applications, we also report the recognition results considering partial matches (when at least 5 or 6 of the 7 characters are correctly recognized) as they may be useful in narrowing down the list of candidate license plates by incorporating additional information such as the vehicle's make and model.

The results of the license plate recognition experiment are shown in Table I. The table demonstrates the recognition accuracy of HR and LR license plate images degraded by bicubic downsampling and recursive Gaussian noise. The difficulty of the task can be seen from the SSIM score, which ranges from 0 to 0.10, as illustrated in Fig. 5, where the license plate characters are barely distinguishable.

TABLE I
RECOGNITION RATES (%) ACHIEVED IN OUR EXPERIMENTS. "ALL" REFERS TO LICENSE PLATES WHERE ALL CHARACTERS WERE RECOGNIZED CORRECTLY; ≥ 6 AND ≥ 5 REFER TO LICENSE PLATES WHERE AT LEAST 6 OR 5 CHARACTERS WERE RECOGNIZED CORRECTLY, RESPECTIVELY.

	RodoSol-ALPR			PKU		
	All	≥ 6	≥ 5	All	≥ 6	≥ 5
OCR [17] — no super-resolution						
HR	96.6	98.6	99.0	99.4	99.9	99.9
LR	0.8	4.6	12.7	0.0	0.0	0.0
OCR [17] — with super-resolution						
Proposed	**39.0**	**59.9**	**74.2**	**72.0**	**90.3**	**97.3**
Nascimento et al. [39]	10.5	25.4	42.2	35.5	65.3	82.5
Mehri et al. [16]	1.45	7.0	17.4	22.5	49.2	70.6
Average PSNR (dB) and SSIM						
		PSNR	SSIM		PSNR	SSIM
Proposed		21.2	**0.59**		**18.3**	**0.61**
Nascimento et al. [39]		**21.3**	0.52		18.1	0.54
Mehri et al. [16]		16.8	0.38		16.4	0.41

The proposed super-resolution network outperformed the baseline models [16], [39] (see the second section of Table I). The multi-task OCR model [17] demonstrated remarkable improvement when applied to images reconstructed by our super-resolution approach in both datasets, particularly in the PKU dataset, with a 14.8% higher recognition rate compared to the method proposed in our preliminary method [39] and a 26.7% higher accuracy compared to MPRNet [16] for license plates with more than five correct characters.

For completeness, we detail in Table I the PSNR and SSIM obtained by each approach. Similar to what was observed in [34], [35], [38], the PSNR metric seems inappropriate for this particular application, as our approach and the one proposed in [39] reached comparable values, despite ours leading to significantly better results achieved by the OCR model. The SSIM metric, on the other hand, seems to better represent the quality of reconstruction of license plate images, as the proposed method achieved considerably better SSIM values in both datasets.

The improved OCR network performance can be attributed to the effective extraction of textural and structural information enabled by the proposed GP unit, along with the optimized channel scaling and reorganization facilitated by the CA and POS units, which utilize pyramid and PS layers.

The variation in accuracy between the two datasets can be attributed to the diversity present in the RodoSol-ALPR dataset, which includes a range of layouts, lighting conditions, and character fonts, while the PKU dataset largely comprises license plates with a uniform layout, with less variation in the conditions under which the images were collected.

Furthermore, the visual comparison of the generated super-resolution images using our technique and the baseline methods [16], [39] supports the results of the license plate recognition experiments. Fig. 7 and Fig. 8 present four pairs of LR and corresponding super-resolution images, along with the original HR image for reference. These images clearly demonstrate that our proposed approach outperforms its preliminary version [39] and MPRNet [16] in terms of perceptual quality.

Fig. 7. Typical examples of the images generated by the proposed approach and baselines in the RodoSol-ALPR dataset [30]. GT = ground truth.

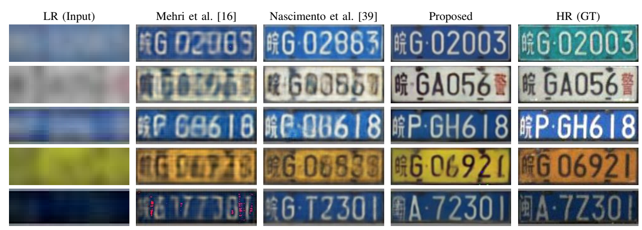


Fig. 8. Representative samples of the images generated by the proposed approach and baselines in the PKU dataset [40]. GT = ground truth.

Common issues observed in MPRNet [16] images include blurriness, where character edges blend into the license plate background, resulting in artifacts. This blurriness can also cause multiple characters' edges to blend together, leading to visual distortions. In contrast, the architecture proposed in [39] successfully reconstructs characters but introduces strong undulations, making them appear as part of the license plate background in some cases (as seen in the first row of Fig. 7). Our proposed model, however, consistently generates clear character edges, accurately reconstructs the original font, and avoids missing characters or incomplete lines.

When uncertain about character reconstruction, our model tends to generate characters most congruent with the LR input, as seen in the last row of Fig. 7 and Fig. 8 (e.g., "3" reconstructed as "J" and "Z" as "2"). Incorporating a lexicon or vocabulary could address this issue by guiding the network to recognize character types based on specific layouts.

Furthermore, the network tends to generate similar background colors for different images, as observed in the second row of Fig. 7 and the first row of Fig. 8. However, our analysis indicates that this has minimal impact on the achieved recognition results.

Finally, it is noteworthy that our model exhibits superior adaptability and ease of training when compared to approaches relying on generative adversarial networks, which often present instability and fall into mode collapse [45], [46]. The attention-based design allows for more straightforward training, making it accessible to a broader range of practitioners and researchers in the field. Moreover, although we did not conduct specific experiments related to execution time, we anticipate that our approach delivers enhanced efficiency when contrasted with architectures based on diffusion models, which are known for being computationally expensive [6], [46].

C. Ablation Study

As our approach integrates multiple concepts into a single architecture, we conducted an ablation study to validate the contribution of each incorporated unit to the results obtained.

Four baselines were established. The first baseline replaced the autoencoder with a 5×5 DConv layer for shallow feature extraction [16]. The second baseline removed the TFAM module and adjusted the output of the previous layer to match the input shape of the following layers. The third baseline replaced the PS and PU layers with transposed and strided convolution layers, respectively, as they are analogous [22]. Finally, in the fourth baseline, the perceptual loss was replaced by MSE, which is commonly used in super-resolution research [3], [4]. Table II presents the results.

TABLE II
RECOGNITION RATES (%) ACHIEVED IN THE ABLATION STUDY. "ALL" DENOTES LICENSE PLATES WHERE ALL CHARACTERS WERE RECOGNIZED CORRECTLY, ≥ 6 AND ≥ 5 REFER TO LICENSE PLATES WHERE AT LEAST 6 OR 5 CHARACTERS WERE RECOGNIZED CORRECTLY, RESPECTIVELY.

Approach	RodoSol-ALPR			PKU		
	All	≥ 6	≥ 5	All	≥ 6	≥ 5
Proposed (w/o autoencoder)	32.7	55.0	70.1	**73.8**	90.2	96.6
Proposed (w/o TFAM)	33.3	55.0	69.6	73.1	90.1	96.6
Proposed (w/o PS and PU layers)	34.3	54.8	68.5	70.4	89.9	96.7
Proposed (w/o perceptual loss)	35.6	57.3	71.9	72.4	**91.4**	97.1
Proposed	**39.0**	**59.9**	**74.2**	72.0	90.3	**97.3**

The experiments on the RodoSol-ALPR dataset showed that each component of the proposed system significantly contributed to its performance. The complete system achieved a recognition rate of 39.0%, while the best version without one component achieved 35.6%. Removing the autoencoder unit had the most detrimental effect, resulting in a recognition rate of 32.7%, as it plays a crucial role in extracting shallow features and guiding the network's reconstruction process.

On the other hand, the recognition rates on the PKU dataset were primarily improved by incorporating the PS and PU layers. It appears that the other units were not necessary for this dataset, which contains less complex images compared to RodoSol-ALPR. This might explain why some authors focused their ablation studies exclusively on the largest and most diverse dataset used in their experiments [13], [42], [47].

V. CONCLUSIONS

This work proposes a new super-resolution approach to improve the recognition of low-resolution license plates. Our method builds upon MPRNet [16] and the architecture proposed in our previous work [39] by incorporating subpixel-convolution layers (PS and PU) in combination with a PLTFAM. We also introduce a novel perceptual loss that combines features extracted from an OCR model with L1 loss to reconstruct characters with the most relevant characteristics, while also incorporating MSE to enhance overall image quality.

Our approach capitalizes on both structural and textural features by using the PS and PU layers for custom scale operations, rather than relying on conventional translational invariance and interpolation techniques. An autoencoder with PS and PU layers was integrated to extract shallow features and generate an attention mask that is added to the original input. The output of the autoencoder is processed by a RDB to identify regions of interest for reconstruction, optimizing computational resources and producing super-resolution images that emphasize relevant information.

Our method achieved superior recognition rates compared to baselines on publicly available datasets from Brazil and mainland China. More specifically, for the RodoSol-ALPR dataset, our method led to a recognition rate of 39.0% being achieved by the OCR model, while the methods proposed in [39] and [16] led to recognition rates of 31.3% and 4.0%, respectively. Similarly, for the PKU dataset, our approach outperformed both baselines, with the OCR model reaching a recognition rate of 72.0%, compared to 35.5% and 22.5% for [39] and [16], respectively. We have made available all datasets used in our experiments (i.e., the LR–HR image pairs), as well as the source code, in order to encourage further research and development in the field of license plate recognition super-resolution.

Future plans include integrating a lexicon or vocabulary into the network's learning process to handle different character types on specific layouts of license plates. Additionally, we aim to create a large-scale dataset for license plate super-resolution collected from thousands of videos, enabling evaluation in real-world scenarios and the development of novel techniques.

Finally, this work generated the following publications:

- A preliminary version of the proposed method was published at the *2022 Conference on Graphics, Patterns and Images (SIBGRAPI)* **[39]**;
- The proposed super-resolution method was recently published in the *Computers & Graphics* journal **[48]**.

ACKNOWLEDGMENTS

This work was supported in part by the Coordination for the Improvement of Higher Education Personnel (CAPES) (*Programa de Cooperação Acadêmica em Segurança Pública e Ciências Forenses # 88881.516265/2020-01*), and in part by the National Council for Scientific and Technological Development (CNPq) (# 308879/2020-1). We gratefully acknowledge the support of NVIDIA Corporation with the donation of the Quadro RTX 8000 GPU used for this research.

REFERENCES

[1] V. Nascimento, "Super-resolution towards license plate recognition," Master's thesis, Federal University of Paraná (UFPR), 2023.

[2] L. Yue, H. Shen, J. Li, Q. Yuan, H. Zhang, and L. Zhang, "Image super-resolution: The techniques, applications, and future," *Signal Processing*, vol. 128, pp. 389–408, 2016.

[3] A. Liu, Y. Liu, J. Gu, Y. Qiao, and C. Dong, "Blind image super-resolution: A survey and beyond," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 5, pp. 5461–5480, 2023.

[4] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 10, pp. 3365–3387, 2021.

[5] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 43, no. 7, pp. 2480–2495, 2021.

[6] M. Santos *et al.*, "Face super-resolution using stochastic differential equations," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2022, pp. 216–221.

[7] G. R. Gonçalves *et al.*, "Multi-task learning for low-resolution license plate recognition," in *Iberoamerican Congress on Pattern Recognition (CIARP)*, Oct 2019, pp. 251–261.

[8] A. Maier *et al.*, "Reliability scoring for the recognition of degraded license plates," in *IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, 2022, pp. 1–8.

[9] D. Moussa *et al.*, "Forensic license plate recognition with compression-informed transformers," in *IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 406–410.

[10] R. Laroca, L. A. Zanlorensi, G. R. Gonçalves, E. Todt, W. R. Schwartz, and D. Menotti, "An efficient and layout-independent automatic license plate recognition system based on the YOLO detector," *IET Intelligent Transport Systems*, vol. 15, no. 4, pp. 483–503, 2021.

[11] Y. Gong *et al.*, "Unified Chinese license plate detection and recognition with high efficiency," *Journal of Visual Communication and Image Representation*, vol. 86, p. 103541, 2022.

[12] S. M. Silva and C. R. Jung, "A flexible approach for automatic license plate recognition in unconstrained scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 6, pp. 5693–5703, 2022.

[13] Y. Wang *et al.*, "Rethinking and designing a high-performing automatic license plate recognition approach," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 7, pp. 8868–8880, 2022.

[14] R. Laroca, L. A. Zanlorensi, V. Estevam, R. Minetto, and D. Menotti, "Leveraging model fusion for improved license plate recognition," in *Iberoamerican Congress on Pattern Recognition*, Nov 2023, pp. 1–15.

[15] A. Lucas *et al.*, "Generative adversarial networks and perceptual losses for video super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3312–3327, 2019.

[16] A. Mehri, P. B. Ardakani, and A. D. Sappa, "MPRNet: Multi-path residual network for lightweight image super resolution," in *IEEE Winter Conference on Applications of Computer Vision*, 2021, pp. 2703–2712.

[17] G. R. Gonçalves *et al.*, "Real-time automatic license plate recognition through deep multi-task networks," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2018, pp. 110–117.

[18] 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, 2016.

[19] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 370–378.

[20] Y. Chen and T. Pock, "Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration," *IEEE Trans. on Pattern Analysis and Machine Intel.*, vol. 39, pp. 1256–1272, 2017.

[21] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 391–407.

[22] W. Shi *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[23] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 294–310.

[24] T. Dai *et al.*, "Second-order attention network for single image super-resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 11 057–11 066.

[25] Y. Huang, J. Li, X. Gao, Y. Hu, and W. Lu, "Interpretable detail-fidelity attention network for single image super-resolution," *IEEE Transactions on Image Processing*, vol. 30, pp. 2325–2339, 2021.

[26] X. He *et al.*, "Ode-inspired network design for single image super-resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019, pp. 1732–1741.

[27] X. Luo, Y. Xie, Y. Zhang, Y. Qu, C. Li, and Y. Fu, "LatticeNet: Towards lightweight image super-resolution with lattice block," in *European Conference on Computer Vision (ECCV)*, 2020, pp. 272–289.

[28] A. Muqeet, J. Hwang, S. Yang, J. Kang, Y. Kim, and S.-H. Bae, "Multi-attention based ultra lightweight image super-resolution," in *European Conference on Computer Vision Workshops*, 2020, pp. 103–118.

[29] Y. Zhang, Y. Huang, K. Wang, G. Qi, and J. Zhu, "Single image super-resolution reconstruction with preservation of structure and texture details," *Mathematics*, vol. 11, p. 216, 01 2023.

[30] R. Laroca, E. V. Cardoso, D. R. Lucio, V. Estevam, and D. Menotti, "On the cross-dataset generalization in license plate recognition," in *International Conference on Computer Vision Theory and Applications (VISAPP)*, Feb 2022, pp. 166–178.

[31] C.-K. Liang, L.-W. Chang, and H. H. Chen, "Analysis and compensation of rolling shutter effect," *IEEE Transactions on Image Processing*, vol. 17, no. 8, pp. 1323–1330, 2008.

[32] K. V. Suresh, G. M. Kumar, and A. N. Rajagopalan, "Superresolution of license plates in real traffic videos," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 321–331, 2007.

[33] J. Yuan, S.-D. Du, and X. Zhu, "Fast super-resolution for license plate image reconstruction," in *International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1–4.

[34] M. Lin, L. Liu, F. Wang, J. Li, and J. Pan, "License plate image reconstruction based on generative adversarial networks," *Remote Sensing*, vol. 13, no. 15, p. 3018, 2021.

[35] A. Hamdi, Y. K. Chan, and V. C. Koo, "A new image enhancement and super resolution technique for license plate recognition," *Heliyon*, vol. 7, no. 11, p. e08341, 2021.

[36] S. Lee, J.-H. Kim, and J.-P. Heo, "Super-resolution of license plate images via character-based perceptual loss," in *IEEE International Conference on Big Data and Smart Computing*, 2020, pp. 560–563.

[37] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision (ECCV)*, 2016, pp. 694–711.

[38] R. Zhang *et al.*, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 586–595.

[39] V. Nascimento, R. Laroca, J. A. Lambert, W. R. Schwartz, and D. Menotti, "Combining attention module and pixel shuffle for license plate super-resolution," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2022, pp. 228–233.

[40] Y. Yuan, W. Zou, Y. Zhao, X. Wang, X. Hu, and N. Komodakis, "A robust and efficient approach to license plate detection," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1102–1114, 2017.

[41] R. Laroca, M. Santos, V. Estevam, E. Luz, and D. Menotti, "A first look at dataset bias in license plate recognition," in *Conference on Graphics, Patterns and Images (SIBGRAPI)*, Oct 2022, pp. 234–239.

[42] L. Zhang *et al.*, "A robust attentional framework for license plate recognition in the wild," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 11, pp. 6967–6976, 2021.

[43] R. Laroca *et al.*, "Do we train on test data? The impact of near-duplicates on license plate recognition," in *International Joint Conference on Neural Networks (IJCNN)*, June 2023, pp. 1–8.

[44] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.

[45] W. Li, L. Fan, Z. Wang, C. Ma, and X. Cui, "Tackling mode collapse in multi-generator GANs with orthogonal vectors," *Pattern Recognition*, vol. 110, p. 107646, 2021.

[46] C. Saharia, J. Ho, W. Chan, T. Salimans, and M. Norouzi, "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713–4726, 2023.

[47] S. Qin and S. Liu, "Towards end-to-end car license plate location and recognition in unconstrained scenarios," *Neural Computing and Applications*, vol. 34, p. 21551–21566, 2022.

[48] V. Nascimento *et al.*, "Super-resolution of license plate images using attention modules and sub-pixel convolution layers," *Computers & Graphics*, vol. 113, pp. 69–76, 2023.