Embedding Similarity Learning for Extreme License Plate Super-Resolution

Abderrezzaq Sendjasni, Mohamed-Chaker Larabi

Embedding Similarity Learning for Extreme License Plate Super-Resolution

Abderrezzaq Sendjasni, Mohamed-Chaker Larabi
CNRS, Univ. Poitiers, XLIM, UMR 7252, France
{*abderrezzaq.sendjasni, chaker.larabi*}*@univ-poitiers.fr*

***Abstract*—Super-resolution (SR) techniques play a crucial role in enhancing the quality of low-resolution images, with significant applications in fields such as security and surveillance, where license plate recognition is critical. This paper focuses on optimizing the super-resolution of license plates using embedding similarity learning. We proposed a novel framework that integrates a Siamese network with a super-resolution model to guide the SR model into enhancing the perceptual quality of reconstructed license plates. By leveraging embedding similarity through Contrastive loss, our approach ensures that the super-resolved images are perceptually and structurally closer to the original ones. The experiments on a synthetic dataset demonstrated that the proposed method outperforms traditional techniques that rely solely on pixel-based loss functions such as MSE. The introduction of embedding similarity loss significantly improves the PSNR and LPIPS metrics, in addition to the optical characters recognition rate.**

***Index Terms*—Super-resolution, License plate, Convolutional neural networks, Embedding similarity, Contrastive learning.**

I. Introduction

Single image super-resolution (SISR) is a well-known research field in computer vision focused on enhancing the spatial resolution and visual fidelity of low-resolution images. Its significance lies in the ability to reconstruct high-resolution details from degraded visual data, thereby improving image quality across diverse applications, including digital photography [1], [2], medical imaging [3], and video surveillance [4]. In particular, SISR has been increasingly applied to enhance license plate (LP) images, where the clarity and legibility of such critical visual data are paramount for effective and reliable automated recognition systems.

License plate recognition (LPR) systems are integral components of modern surveillance, traffic management, and security applications [5]. However, the efficacy of these systems heavily relies on the quality of the captured images [6]. In real-world scenarios, LP images obtained from surveillance cameras or other sources often suffer from visual degradation such as low resolution, motion blur, and noise. An illustration is provided in Fig. 1, where it appears challenging to properly read the LPs. These issues pose greater challenges to LPR systems, compromising their accuracy and reliability. One of the most critical challenges is the distance from where the images are taken, which induces limited pixel resolution. When images are captured from a great distance, the LPs appear smaller within the frame, and the number of pixels representing the plate is significantly reduced. This reduced pixel resolution results in fine details becoming invisible and unrecognizable upon zooming, making it difficult for LPR systems and even human users to accurately interpret and recognize the characters. As a solution, advanced image enhancement techniques, such as SISR, are essential to mitigate these issues by upscaling low-resolution images and preserving crucial details necessary for reliable LPR.

Fig. 1. Example of license plates taken under different conditions, showcasing difficulties to properly read the plates in some cases [7].

Despite recent advances in the field of SISR [8]–[10], LP super-resolution (LPSR) remains a significant challenge. The unique characteristics of LP images, such as small text, complex backgrounds, varying lighting conditions, and diverse fonts [11]–[14], make it difficult for standard SISR models to perform robustly. Traditional SR methods often fail to reconstruct the fine details and sharp edges required for accurately recognizing characters on license plates. Deep learning-based methods, while more robust, still struggle to maintain the delicate balance between enhancing image quality and preserving critical textual information. The importance of achieving high accuracy in LPR systems cannot be overstated, especially in applications such as law enforcement. Misrecognition or failure to recognize a LP can lead to severe consequences, making it of paramount importance to develop SR methods specifically designed to address these challenges.

Compared to the extensive literature on SISR, research specifically focusing on LPSR remains relatively week. Most existing approaches are based on deep learning, mainly due to their ability to use inherent prior knowledge of natural scenes and preserve image details better than traditional methods, such as interpolation-based ones. For instance, bilinear and bicubic interpolations are simple and fast, but often produce blurry images with a loss of fine details. These techniques estimate new pixel values based on linear or cubic inter-

This work is supported by The French Research Funding Agency (ANR), project IMPROVED ANR-22-CE39-0006.

polation of neighboring pixels, which can result in smooth but less detailed outputs. In contrast, deep learning-based methods, such as convolutional neural networks (CNNs) and generative adversarial networks (GANs), have shown superior performances in LPSR [15]. Their ability to learn complex patterns enables them to reconstruct high-resolution images with finer details and higher visual fidelity.

Motivated by the performance of CNNs and GANs, the work in [16] proposed a multi-scale CNN model trained to minimize the mean squared error (MSE) between the high-resolution (HR) and super-resolved (SR) license plates. The work in [17] proposed to train a GAN architecture with a gradient profile prior [18] to improve the character boundaries to emphasize the contrast between the characters and the background. Furthermore, the work in [15] adopted a similar architecture with GANs. Inspired by the SRGAN model [2], the authors trained a GAN with an OCR-based loss function, computing recognition errors between HR and SR license plates. Additionally, it uses perceptual loss based on VGG-19 [19], and adversarial loss to improve both the visual and recognition accuracy of the SR images. Following the idea of employing OCR as a loss function to guide the learning process, the work in [20] integrated a CNN as a sub-net with a GAN to compute the OCR on the generated image from the generative network of the proposed model. The predicted OCR is compared to the ground truth LP using connectionist temporal classification (CTC) loss. The work in [21] adopted a character-based perceptual loss, where MSE between intermediate feature representations, obtained by means of a character classification model, is used to compute the loss. The work in [22] computes the Levenshtein distance between the predicted characters by an OCR on the HR and SR images as a loss function, combined with MSE and structural similarity index (SSIM) to improve the overall performance.

In summary, current research in LPSR primarily applies existing SISR models, such as SRGAN, augmented with text recognition guidance. Besides, most frameworks focus on moderate scaling factors like x4 and x8, often neglecting extreme cases such as x16 scaling. This study addresses extreme LPSR scenarios, *i.e.* x16, by proposing a novel training approach based on embedding similarity. Specifically, we introduce RDASRNet, a CNN model featuring residual dense blocks (RDBs) and channel attention mechanisms. In addition, we propose a training strategy to improve the fidelity between HR and SR images by minimizing pixel-level distances using MSE and embedding-level differences via a Siamese network and a Contrastive loss [23]. The feature embedding distance is computed based on the output embedding of the Siamese network [24] of the HR and SR images. By doing so, we ensure that the super-resolved images exhibit high visual fidelity and sharpness and accurately preserve the informative details essential for effective LPR. We validate the experiment on the UKLPD dataset [25], composed of synthetic data of UK license plate images.

II. Proposed Methodology

In this section, we detail the architecture of the proposed framework, including the data preparation and degradation modeling, the structure of RDASRNet, and the training strategy based on Siamese network and embedding similarity learning.

A. Data preparation and degradation modeling

The goal of an SR model is to reconstruct HR images from LR images without prior knowledge of the degradation process. Typically, a degradation model is used, which may include blur, noise, and downsampling to synthesize LR images [10]. This can be mathematically modeled by:

$$\mathbf{I}_{LR} = (\mathbf{I}_{HR} * \mathbf{k}) \downarrow_x + \mathbf{n}, \tag{1}$$

where $\mathbf{I}_{LR}$ and $\mathbf{I}_{HR}$ represent the low-resolution and high-resolution images, respectively. The $*$ denotes the convolution operation with a blur kernel $\mathbf{k}$, $\downarrow_x$ indicates downsampling by a factor of x, and $\mathbf{n}$ represents additive noise.

Blur, noise, and downsampling are specifically used in super-resolution tasks because they represent common real-world factors that degrade image quality, including optical imperfections, sensor noise, and environmental conditions [13], [14]. By modeling these specific degradations, SR algorithms can be trained to handle a wide range of real-world scenarios, making them more effective and versatile. Motivated by the fact that real degradation processes are much more diverse and not limited to camera sensor degradations, we incorporate JPEG compression along with the aforementioned degradations into the degradation modeling. By doing so, we account for digital image storage and transmission-related issues, ensuring that the super-resolution model can handle a broader spectrum of real-world image quality issues.

Instead of using license plate images directly, we use localized patches, a common practice in SISR. Each input image $\mathbf{I}_{\text{LP}}$ is divided into non-overlapping patches of size 64×64 pixels. By doing so, we ensure that the model can learn from a diverse set of localized image features. Then, we perform a single-stage degradation process to simulate real-world visual quality issues. Each patch in the dataset undergoes three separate degradation processes: Gaussian Blur (GB), Gaussian Noise (GN), and JPEG Compression. The parameters for these degradations are set to 2 for GB, 0.03 for GN, and 60 for JPEG. Finally, the degraded patches are downsampled by a factor of 16 to simulate extreme loss of resolution:

$$\{\mathbf{P}_{i,j}^{\text{gb}} \downarrow_{16}, \quad \mathbf{P}_{i,j}^{\text{gn}} \downarrow_{16}, \quad \mathbf{P}_{i,j}^{\text{jpeg}} \downarrow_{16}\}. \tag{2}$$

An illustration of the degradation results is provided in Fig. 2, along with the quality scores of the degraded patches as evaluated by the LPIPS [26] model and the difference maps. The LPIPS scores and the associated maps clearly illustrate that each degradation process impacts the perceptual quality of the image patch in distinct ways. GB introduces a moderate perceptual difference compared to the original patch,

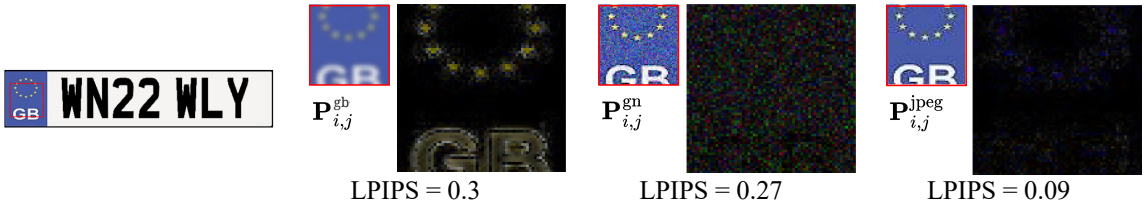

Fig. 2. Illustration of the degradations on a cropped patch (highlighted by the red rectangle), along with its corresponding perceptual difference map and evaluated quality score using LPIPS [26].

whereas GN results in a slightly lower perceptual difference than GB, typically introducing graininess and random pixel variations that affect the image's texture. JPEG compression, using the specified quality factor, results in the least perceptual difference among the three degradations.

B. RDASRNet structure

To achieve high-quality super-resolution, we designed a model based on foundational observations from state-of-the-art SISR techniques, leveraging residual dense blocks (RDBs) [27] and attention mechanism [28].

The model initiates with extracting essential visual features $\mathbf{X} \in \mathbb{R}^{H\times W\times C}$, where H, W, and C denote the height, width, and number of channels, respectively, from the LR input patches. Then a sequence of RDBs is employed to capture intricate hierarchical features effectively. Each RDB iteration refines feature maps through dense connections and residual learning:

$$\mathbf{X}^{(i+1)} = \mathcal{F}_{\text{RDB}}(\mathbf{X}^{(i)}), \tag{3}$$

where $\mathbf{X}^{(i)}$ represents feature maps at the i-th RDB iteration, and $\mathcal{F}_{\text{RDB}}$ encapsulates the operations within the RDB, including multiple convolutional layers and nonlinear activations, performed by rectified linear units (ReLU). Then, to enhance the feature representations and emphasize informative channel features, a channel attention mechanism is integrated. This mechanism computes attention scores as follows:

$$\mathbf{X}_{\text{CA}} = \sigma(\text{FC}(\text{ReLU}(\text{FC}(\text{GAP}(\mathbf{X}))))), \tag{4}$$

where $\mathbf{X}_{\text{CA}}$ represent the refined set of features, σ denotes the sigmoid activation function, and GAP represents global average pooling. This step enables the model to adaptively recalibrate channel-wise feature responses [28], focusing on relevant information, considered paramount for accurate SR.

Subsequently, the model employs dynamic upsampling via multiple 2D transposed convolution operations [29]. Each one incrementally doubles the spatial dimensions of the feature maps, progressively enhancing spatial resolution while preserving essential details. This multi-stage upsampling strategy allows the model to refine and upscale the output patch across several levels, thereby mitigating artifacts and enhancing overall image fidelity:

$$\mathbf{X}_{\text{up}} = \text{ReLU}\left(\text{ConvT}^{L}(\mathbf{X}_{\text{CA}})\right). \tag{5}$$

Given the scale factor x, the upsampling process involves $L = \log_2(x)$ stages, where each stage doubles the spatial dimensions of the feature maps. ConvT represent the 2D transposed convolution. Finally, the SR patch $\mathbf{P}_{\text{SR}}$ is generated through a final convolutional layer:

$$\mathbf{P}_{\text{SR}} = \text{Conv2d}(\mathbf{X}_{\text{up}}). \tag{6}$$

C. Siamese and Embedding similarity learning

To effectively enhance the LPSR, we designed a loss function that combines pixel-wise loss with embedding similarity learning using a Siamese network architecture [24]. The total loss function $\mathcal{L}_{\text{P-EM}}$ is a weighted sum of the MSE loss $\mathcal{L}_{\text{pixel}}$ and the Contrastive loss $\mathcal{L}_{\text{contrastive}}$, accounting for the embedding similarity loss.

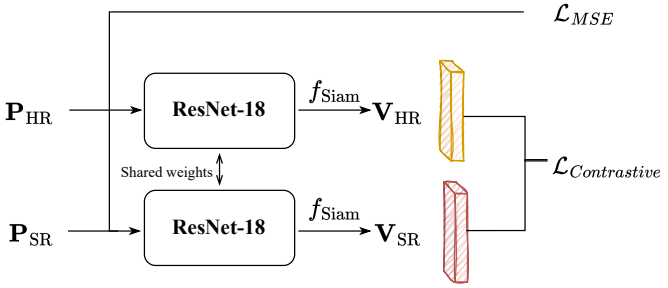

Fig. 3. Illustration of the proposed loss structure for LPSR. The Siamese network extracts embeddings from the SR and HR patches. The contrastive loss function penalizes the distance between the resulting embeddings.

Pixel-wise loss: This part of the loss is computed using the MSE between $\mathbf{P}_{\text{SR}}$ and $\mathbf{P}_{\text{HR}}$, and can be obtained as:

$$\mathcal{L}_{\text{pixel}} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{P}^{(i)}_{\text{SR}} - \mathbf{P}^{(i)}_{\text{HR}})^2, \tag{7}$$

where N represents the number of pixels in the image patch, and $\mathbf{P}^{(i)}_{\text{SR}}$ and $\mathbf{P}^{(i)}_{\text{HR}}$ denote the pixel values of the SR and HR patches, respectively, at the i-th position. This loss measures the average of the squared differences between corresponding pixel values.

Contrastive loss: To ensure that the super-resolved patches retain the key features of the HR patches, we employ a Siamese network [24] to extract embeddings from both the predicted SR and HR patches. A Siamese neural network

is a class of neural network architectures that contain two identical sub-networks. The same configuration with the same parameters and weights is shared between them. Parameter updating is mirrored across both sub-networks, to ensure that the embeddings generated are directly comparable.

The Siamese network is based on a pre-trained ResNet-18 model, with the last fully connected layer replaced to produce embeddings of size 128. Given an input pair $(\mathbf{P}_{\text{SR}}, \mathbf{P}_{\text{HR}})$ as depicted in Fig. 3, the Siamese network outputs two embeddings $\mathbf{V}_{\text{SR}}$ and $\mathbf{V}_{\text{HR}}$:

$$\mathbf{V}_{\text{SR}}, \mathbf{V}_{\text{HR}} = f_{\text{Siam}}(\mathbf{P}_{\text{SR}}, \mathbf{P}_{\text{HR}}; \theta_{\text{Siam}}), \tag{8}$$

where θ_{Siam} represents the parameters of the Siamese network.

The goal of the Siamese network is to predict similar embeddings for the SR and HR patches, reflecting their similarity. To achieve this, we use the Contrastive loss function $\mathcal{L}_{\text{Contrastive}}$. The latter encourages the embeddings of the SR and the HR patches to be similar. Besides, the RDASRNet is supposed to generate a similar embedding to the HR patch, where both $\mathbf{P}_{\text{SR}}$ and $\mathbf{P}_{\text{HR}}$ are considered similar. Therefore, the original Contrastive loss [23] is simplified to:

$$\mathcal{L}_{\text{Contrastive}} = \max(m - D, 0)^2. \tag{9}$$

In this context:

- D represents the Euclidean distance between the embeddings of the SR and HR patches. It can be obtained by:

$$D = \|\mathbf{v}_{\text{SR}} - \mathbf{v}_{\text{HR}}\|_2. \tag{10}$$

- m is the margin. It is a hyperparameter that defines a threshold distance between the embeddings, and is set to 2 in this study.
- The squaring operation $(m - D)^2$ penalizes larger deviations more heavily, enforcing a stronger push towards reducing the distance D when it is less than m.

The emphasis on embedding similarity loss is motivated by the use of LPSR images in optical character recognition (OCR), automatic number plate recognition (ANPR), and vehicle identification. Ensuring similar embeddings guarantees that the SR images maintain the distinctive features found in HR images. This alignment enhances the performance of OCR, ANPR, and vehicle identification systems, where accurate feature preservation is critical for reliable recognition and identification.

Total Loss: The final loss $\mathcal{L}_{\text{P-EM}}$ is a weighted sum of the pixel-wise loss and the Contrastive loss, defined as:

$$\mathcal{L}_{\text{P-EM}} = \alpha \mathcal{L}\text{pixel} + \beta \mathcal{L}_{\text{Contrastive}}, \tag{11}$$

where α and β are regularization weights that balance the contribution of each loss component.

III. Experiments

A. Datasets and Implementation Details

Dataset: For this study, we use the UKLPD synthetic dataset [25] to validate the proposed framework. The dataset contains synthetic images of UK license plates that conform to standard letter sizes and spacing prescribed by the driver and vehicle licensing agency. In total, 24,000 images are provided, evenly split between white and yellow backgrounds.

For training and evaluation, the dataset is partitioned into 90% for training and 10% for validation sets. This splitting ensures the model has sufficient data for learning, while maintaining a separate subset for unbiased performance evaluation. Each image is segmented into patches of size 64×64 pixels, and each patch undergoes the degradation process detailed in Section II-A, resulting in over 300,000 training samples.

Implementation Details: The RDASRNet is implemented using the PyTorch library [30] and trained on a server equipped with an Intel Xeon Silver 4208 2.1GHz CPU, 192GB of RAM, and an Nvidia Tesla V100S GPU with 32GB of memory. We train the model for 200 epochs, a batch size of 128, and the Adam optimizer [31] with a learning rate of 1×10^{-4}.

Evaluation criteria: Two important aspects should be evaluated for LPSR: visual quality and character recognition rates. The visual quality assesses how perceptually similar the super-resolved images are to the ground truth high-resolution images. For this, we use the PSNR metric, with the highest score the better, and the LPIPS model with lower scores the better. As for the character recognition rates, they are used to evaluate the effectiveness of the super-resolution model in enhancing the legibility of characters on license plates (LPs). This involves applying OCR techniques to the SR images. For this, we use EasyOCR [32], which provides confidence scores on the character recognition.

B. Results and discussion

With the aim to evaluate the effectiveness of the proposed loss function, $\mathcal{L}_{\text{P-EM}}$, which combines pixel and embedding similarity losses, we conducted a comparative analysis against the commonly used $\mathcal{L}_{\text{MSE}}$. Fig. 4 presents a box plot of PSNR and LPIPS metrics for each degradation separately, illustrating the distribution of median, minimum, and maximum scores. This visualization highlights the variability of the model performance across different loss configurations. For the proposed $\mathcal{L}_{\text{P-EM}}$, we systematically varied the regularization parameters α and β to examine their impact on the balance between pixel accuracy and embedding similarity. This analysis allows us to identify the contributions of each component of the loss function to the overall image quality, providing insights into the optimal parameter settings for better performances.

From the box plots, the performances actively demonstrate that $\mathcal{L}_{\text{P-EM}}$ consistently outperforms $\mathcal{L}_{\text{MSE}}$ in both PSNR and LPIPS metrics. This indicates that incorporating embedding similarity into the loss function significantly enhances the model's ability to generate visually and perceptually accurate high-resolution images. The variations in the α/β ratio further

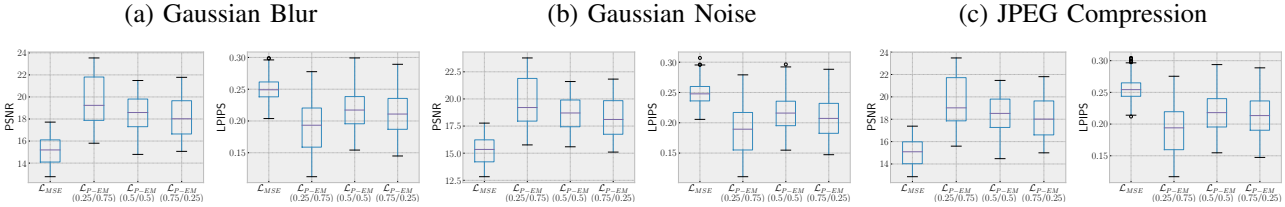

Fig. 4. Performances comparison of the proposed $\mathcal{L}_{\text{P-EM}}$ with different α/β values over the use of traditional $\mathcal{L}_{\text{MSE}}$ on each degradation.

depict the sensitivity of the model to different aspects of image features, suggesting that a higher weight on the embedding similarity loss tends to produce better perceptual quality while maintaining robust pixel-level accuracy. Across the degradations, the performances show stability, depicting the model's generalization thanks to the training strategy.

With regards to OCR's ability to effectively recognize text on the generated license plates, the EasyOCR metric provides a confidence score for the predicted text. This confidence score increases with sharper details and robust representative features. We compute the OCR confidence scores across various configurations to assess its impact on text recognition accuracy. The results are shown with a box plot in Fig. 5. Higher confidence scores indicate that the OCR system is more likely to correctly identify the characters on the license plates, which is essential for applications in security and surveillance. By integrating embedding similarity learning, our model significantly enhances the clarity and legibility of the license plates, thereby boosting OCR confidence.

Based on the performances, we observe that the OCR confidence scores improved consistently when using the proposed $\mathcal{L}_{\text{P-EM}}$ loss function compared to the traditional $\mathcal{L}_{\text{MSE}}$. This is valid regardless of the degradation. Specifically, the confidence scores demonstrate that the inclusion of embedding similarity not only preserves the structural integrity of the characters but also maintains high perceptual quality, leading to better recognition results.

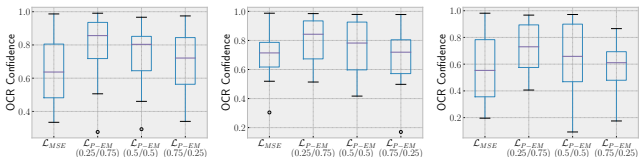

Fig. 5. OCR confidence intervals comparison on each degradation.

In addition to the overall performance metrics, Fig. 6 presents a qualitative evaluation using two samples with different background colors. The first row displays the HR image, while the second row shows the reconstructed LR image after downsampling by a factor of x16. Visually, the license plate becomes completely unreadable following this extreme downsampling. The remaining rows illustrate the outputs of the super-resolution model using different configurations of the loss function. The obtained results demonstrate significant efficacy, with SR license plates recognizable, even though

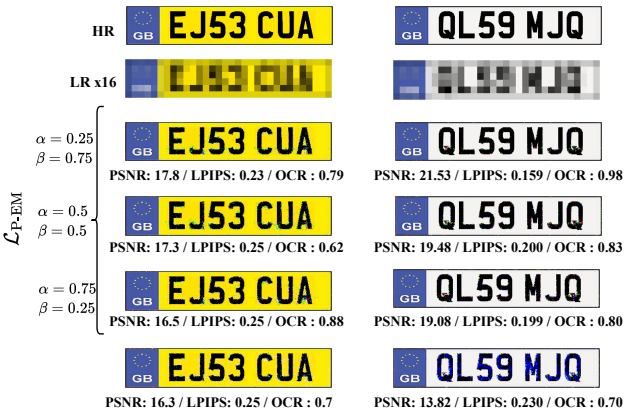

Fig. 6. Qualitative evaluation of license plate samples from the UKLPD testing set with the JPEG compression and x16 downsampling.

some noise and color deformation are generated. The quality scores are mostly impacted by the noise, resulting in low PSNR values, particularly for the $\mathcal{L}_{\text{MSE}}$. This depicts a strong correlation with the OCR confidences. The latter decreases from 0.99 to 0.80 with increased weight on the $\mathcal{L}_{\text{pixel}}$ component and further drops to 0.67 when training exclusively with $\mathcal{L}_{\text{MSE}}$ on the sample with a white background. However, on the sample with a yellow background, $\mathcal{L}_{\text{MSE}}$ performed better, though still inferior to the proposed $\mathcal{L}_{\text{P-EM}}$. These findings align with the performance trends in Fig. 4, indicating that prioritizing feature embedding similarity in SR image generation enhances OCR system accuracy more than solely emphasizing pixel fidelity. Besides, the background-character contrast appears to significantly impact the performances. As MSE focuses on pixel-level differences, it might struggle with high-contrast edges and fine details, resulting in relatively poorer performance for white backgrounds. This contrast sensitivity suggests that emphasizing feature embedding similarity is better across varying background conditions, thereby ensuring more consistent OCR accuracy.

To further evaluate the efficacy of the proposed training strategy, a comparison with interpolation-based methods, including bicubic and bilinear interpolation, is provided in Table III-B. This comparison is performed on the sample with the yellow background from Fig. 6. The performances in terms of PSNR and LPIPS for each degradation are reported. As it can be seen, the proposed framework consistently and significantly outperforms the interpolation-based methods across

all degradation types. Approx. 7db is gained over bicubic and bilinear in terms of PSNR, and approx. 0.4 is gained with LPIPS. This demonstrates the superior ability of the proposed method to reconstruct high-quality and perceptually accurate super-resolution images, confirming its effectiveness in handling different degradation scenarios.

TABLE I
PERFORMANCE COMPARISON WITH INTERPOLATION-BASED METHODS IN TERMS OF PSNR (BEST IN RED) AND LPIPS (BEST IN BLUE).

Deg.	Metric	Bicubic	Bilinear	RDASRNet $\mathcal{L}_{\text{MSE}}$	RDASRNet $\mathcal{L}_{\text{P-EM}_{(0.25/0.75)}}$
GB	PSNR	9.80	10.59	16.49	17.79
	LIPIS	0.7244	0.6617	0.2348	0.2340
GN	PSNR	10.21	10.53	16.60	17.89
	LIPIS	0.7085	0.6607	0.2423	0.2265
JPEG	PSNR	9.74	10.57	16.3521	17.57
	LIPIS	0.7234	0.6537	0.2471	0.2335

IV. CONCLUSION

In this study, we developed and evaluated an advanced approach for LPSR using a novel loss function that combines pixel and embedding similarity losses. Our findings demonstrate that this approach significantly enhances the perceptual quality of SR images compared to solely relying on pixel-to-pixel fidelity loss. Through comprehensive quantitative and qualitative evaluations, we observed substantial improvements in PSNR and LPIPS scores, indicating superior visual fidelity and perceptual quality generation. Moreover, the proposed method shows promising results in enhancing OCR confidence, crucial for forensic applications. Its effectiveness underscores the importance of incorporating feature embedding similarity alongside pixel-level fidelity metrics in super-resolution tasks. For future work, we plan to explore multi-stage degradation modeling, incorporating geometric distortions and other complex degradations. Besides, we aim to leverage the insights from the used synthetic dataset, *i.e.* UKLPD, to enhance the accuracy on real-world data through fine-tuning and transfer learning techniques.

REFERENCES

[1] Xintao Wang, Ke Yu, Shixiang Wu, et al. ESRGAN: Enhanced super-resolution generative adversarial networks. In *IEEE ECCVw*, pages 0–0, 2018.

[2] Christian Ledig, Lucas Theis, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE CVPR*, July 2017.

[3] Eric Van Reeth, Ivan WK Tham, Cher Heng Tan, and Chueh Loo Poh. Super-resolution in magnetic resonance imaging: a review. *CMR*, 40(6):306–325, 2012.

[4] Prayushi Mathur, Ashish Kumar Singh, Syed Azeemuddin, Jayram Adoni, and Prasad Adireddy. A real-time super-resolution for surveillance thermal cameras using optimized pipeline on embedded edge device. In *IEEE AVSS*, pages 1–7, 2021.

[5] Gee-Sern Hsu, Jiun-Chang Chen, and Yu-Zu Chung. Application-oriented license plate recognition. *IEEE TVT*, 62(2):552–561, 2013.

[6] Valfride Nascimento, Rayson Laroca, Jorge de A. Lambert, William Robson Schwartz, and David Menotti. Super-resolution of license plate images using attention modules and sub-pixel convolution layers. *Computers and Graphics*, 113:69–76, 2023.

[7] R. Laroca, E. V. Cardoso, D. R. Lucio, V. Estevam, and D. Menotti. On the cross-dataset generalization in license plate recognition. In *ICVISAPP*, pages 166–178, Feb 2022.

[8] Brian B. Moser, Federico Raue, Stanislav Frolov, Sebastian Palacio, Jörn Hees, and Andreas Dengel. Hitchhiker's guide to super-resolution: Introduction and recent advances. *IEEE TPAMI*, 45(8):9862–9882, 2023.

[9] Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Blind image super-resolution: A survey and beyond. *IEEE TPAMI*, 45(5):5461–5480, 2023.

[10] Yawei Li, Yulun Zhang, Radu Timofte, Luc Van Gool, Lei Yu, et al. Ntire 2023 challenge on efficient super-resolution: Methods and results. In *IEEE/CVF CVPR*, pages 1921–1959, 2023.

[11] Gabriel Resende Gonçalves, Matheus Alves Diniz, Rayson Laroca, David Menotti, and William Robson Schwartz. Multi-task learning for low-resolution license plate recognition. In *Progress in PRIACVA*, pages 251–261. Springer, 2019.

[12] Anatol Maier, Denise Moussa, Andreas Spruck, Jürgen Seiler, and Christian Riess. Reliability scoring for the recognition of degraded license plates. In *IEEE AVSS*, pages 1–8, 2022.

[13] Denise Moussa, Anatol Maier, Andreas Spruck, Jürgen Seiler, and Christian Riess. Forensic license plate recognition with compression-informed transformers. In *IEEE ICIP*, pages 406–410, 2022.

[14] Yanxiang Gong, Linjie Deng, et al. Unified chinese license plate detection and recognition with high efficiency. *JVCIR*, 86:103541, 2022.

[15] Yuecheng Pan, Jin Tang, and Tardi Tjahjadi. Lpsrgan: Generative adversarial networks for super-resolution of license plate image. *Neurocomputing*, page 127426, 2024.

[16] Yang Yang, Ping Bi, and Ying Liu. License plate image super-resolution based on convolutional neural network. In *IEEE ICIVC*, pages 723–727, 2018.

[17] Yu Lu, Yu Gu, and Bi Wang. License plate recognition in wild with super-resolution. In *ICNNICE)*, pages 523–526, 2023.

[18] Jian Sun, Jian Sun, Zongben Xu, and Heung-Yeung Shum. Gradient profile prior and its applications in image super-resolution and enhancement. *IEEE TIP*, 20(6):1529–1542, 2011.

[19] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *IEEE/CVF ICCV*, pages 1905–1914, October 2021.

[20] Zuzana Bílková and Michal Hradiš. Perceptual license plate super-resolution with ctc loss. *EI*, 32:1–5, 2020.

[21] Seyun Lee, Ji-Hwan Kim, and Jae-Pil Heo. Super-resolution of license plate images via character-based perceptual loss. In *IEEE BigComp*, pages 560–563, 2020.

[22] Valfride Nascimento, Rayson Laroca, Jorge de A. Lambert, William Robson Schwartz, and David Menotti. Combining attention module and pixel shuffle for license plate super-resolution. In *SIBGRAPI*, volume 1, pages 228–233, 2022.

[23] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *IEEE CVPR)*, volume 2, pages 1735–1742, 2006.

[24] Gregory Koch, Richard Zemel, Ruslan Salakhutdinov, et al. Siamese neural networks for one-shot image recognition. In *ICML deep learning workshop*, volume 2, 2015.

[25] Saad Bin Munir. Uk licence plate synthetic images. *University of Central Lancachir*, 2021.

[26] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE CVPR*, pages 586–595, 2018.

[27] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *IEEE CVPR*, pages 2472–2481, 2018.

[28] Haoyu Chen, Jinjin Gu, and Zhi Zhang. Attention in attention network for image super-resolution. *arXiv preprint arXiv:2104.09497*, 2021.

[29] Hongyang Gao, Hao Yuan, Zhengyang Wang, and Shuiwang Ji. Pixel transposed convolutional networks. *IEEE TPAMI*, 42(5):1218–1227, 2020.

[30] Jason Ansel, Edward Yang, Horace He, Natalia Gimelshein, et al. Pytorch 2: Faster machine learning through dynamic python bytecode transformation and graph compilation. In *ACM ICASPL*, page 929–947, 2024.

[31] D. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[32] Rakpong Kittinaradorn, Wisuttida Wichitwong, Nart Tlisha, et al. cwitt-wer/easyocr: Easyocr, July 2022.