MF-LPR²: Multi-Frame License Plate Image Restoration and Recognition using Optical Flow⋆

Kihyun Na[a], Junseok Oh[a,1], Youngkwan Cho[a], Bumjin Kim[a], Sungmin Cho[a], Jinyoung Choi[a] and Injung Kim[a,*]

[a]Department of Computer Science and Electrical Engineering (CSEE), Handong Global University, 558 Handong-ro, Buk-gu, Pohang, 37554, Gyeongsangbuk, Republic of Korea

ARTICLE INFO

Keywords:
license plate recognition
license plate image restoration
license plate image dataset
super resolution
optical flow

ABSTRACT

License plate recognition (LPR) is important for traffic law enforcement, crime investigation, and surveillance. However, license plate areas in dash cam images often suffer from low resolution, motion blur, and glare, which make accurate recognition challenging. Existing generative models that rely on pretrained priors cannot reliably restore such poor-quality images, frequently introducing severe artifacts and distortions. To address this issue, we propose a novel multi-frame license plate restoration and recognition framework, MF-LPR², which addresses ambiguities in poor-quality images by aligning and aggregating neighboring frames instead of relying on pretrained knowledge. To achieve accurate frame alignment, we employ a state-of-the-art optical flow estimator in conjunction with carefully designed algorithms that detect and correct erroneous optical flow estimations by leveraging the spatio-temporal consistency inherent in license plate image sequences. Our approach enhances both image quality and recognition accuracy while preserving the evidential content of the input images. In addition, we constructed a novel Realistic LPR (RLPR) dataset to evaluate MF-LPR². The RLPR dataset contains 200 pairs of low-quality license plate image sequences and high-quality pseudo ground-truth images, reflecting the complexities of real-world scenarios. In experiments, MF-LPR² outperformed eight recent restoration models in terms of PSNR, SSIM, and LPIPS by significant margins. In recognition, MF-LPR² achieved an accuracy of 86.44%, outperforming both the best single-frame LPR (14.04%) and the multi-frame LPR (82.55%) among the eleven baseline models. The results of ablation studies confirm that our filtering and refinement algorithms significantly contribute to these improvements.

1. Introduction

License Plate Recognition (LPR) plays a crucial role in various traffic applications, including traffic law enforcement, crime investigation, and surveillance. However, in road video footage, license plates often appear small, even in high-resolution images (e.g., 2560x1440). As a result, the resolutions of the license plate image are often insufficient for recognition (e.g., 88x32). This issue is further compounded by other degradations, such as motion blur, light glare, and inaccurate focus, which makes reliable recognition challenging in real-world scenarios.

License plate image restoration is distinguished from general image restoration, which primarily focuses on visual quality, in that it must enhance legibility while preserving the evidential value of the original footage. For example, the model must not restore the degraded image of the digit '6' in a way that makes it resemble '5' or '8', as such errors compromise the forensic value of the image.

However, most recent restoration methods often fail to preserve the original information of images, introducing artifacts or distortions due to their heavy reliance on prior knowledge acquired from training data.

⋆ Our RLPR dataset can be found in 10.17632/4rs5wpvckz.2

*Corresponding author

kevinna95@gmail.com (K. Na); junseokoh96@gmail.com (J. Oh); dudrhks1009@naver.com (Y. Cho); 21900104@handong.ac.kr (B. Kim); ko041213@gmail.com (S. Cho); jinyoung@handong.ac.kr (J. Choi); ijkim@handong.edu (I. Kim)

ORCID(s): 0009-0001-3827-5371 (K. Na); 0009-0001-5450-8956 (J. Oh); 0009-0004-9821-8688 (Y. Cho); 0009-0009-6331-6938 (B. Kim); 0009-0008-0815-5441 (S. Cho); 0009-0002-6255-2882 (J. Choi); 0000-0003-4439-6097 (I. Kim)

[1]Currently, working at Samsung Electronics.

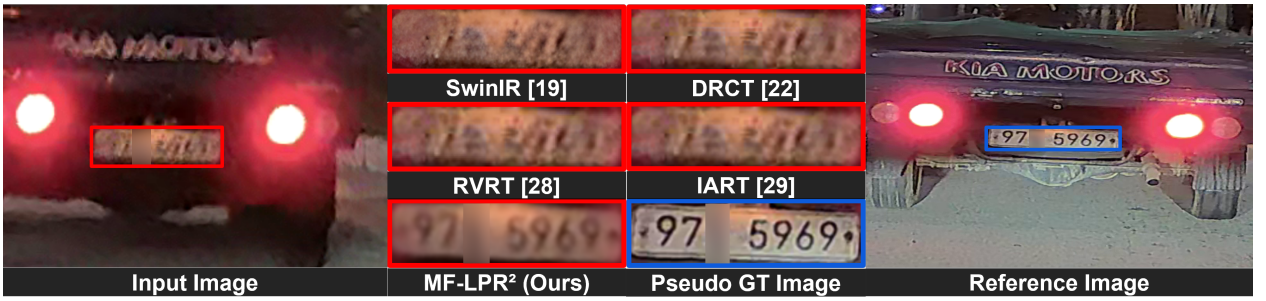

Figure 1: The result of the proposed model (MF-LPR²) compared with two single image super resolution (SISR), SwinIR(10), DRCT(13), and two video restoration models (VSR), RVRT (19) and IART (20). The left column displays the low-quality input image. The middle six images present the results of the five restoration models and as well as the pseudo ground-truth image extracted from the reference image (right column). MF-LPR² showed the best result both preserving the content of the input image and improving visibility.

The advent of deep generative models has led to significant advancements in image restoration, enabling the reconstruction of missing details leveraging pretrained knowledge acquired from large datasets. This approach has been widely applied in tasks such as single-image super-resolution (SISR) (1; 2; 3; 4; 5; 6; 7; 8; 9; 10; 11; 12; 13; 14), multi-frame super-resolution (MFSR) (50), and video super-resolution (VSR) (15; 16; 17; 18; 19; 20), demonstrating substantial improvements in visual quality. In particular, recently developed scene text image super-resolution (STISR) models (21; 22; 23; 24; 25; 26) have shown their effectiveness in enhancing the resolution and readability of low-resolution text images by leveraging the semantic information unique to textual contents.

However, existing generative models often struggle with severely degraded license plate images, as shown in Fig. 1. This challenge is particularly pronounced for images that differ significantly from the characteristics of the training data. The primary reasons for this limitation is as follows: First, the degradation in the input image quality surpasses the range that restoration models can handle. Second, generative models tend to prioritize the enhancement of visual quality by relying heavily on pretrained knowledge, rather than focusing on preserving the content of the input images. Third, restoration models are generally trained on synthetic image pairs consisting of high-resolution images and their artificially degraded counterparts, as collecting paired datasets of low- and high-quality license plate images is resource-intensive and costly. However, the degradations observed in real-world license plate images differ significantly from these artificial degradations.

There are several previous studies on license plate image restoration (42; 43; 44; 45; 50). Most of these studies focus on motion deblurring and the issues related to low resolution and quality have received relatively less attention. While several previous studies have attempted to extend restoration beyond motion deblurring (47; 48; 49), these models were trained on synthetic datasets, limiting their effectiveness in real-world scenarios with diverse degradation patterns.

In this study, our objective is to restore low-quality license plate images in a manner that enhances their legibility while preserving the evidential value by preventing any alteration of the original content. To achieve this goal, we propose a novel multi-frame license plate image restoration and recognition (MF-LPR²) framework. Unlike generative models that rely on prior knowledge obtained from training data, MF-LPR² restores high-quality license plate images by combining the complementary information from multiple low-resolution input frames.

MF-LPR² aligns a sequence of low-quality frames using optical flows and aggregates the aligned frames to produce a high-quality output image. However, while accurate optical flow estimation is critical in this approach, even state-of-the-art optical flow estimators often produce erroneous results for low-quality images, which severely degrades output quality as Fig. 4. To overcome this challenge, we present reliable optical flow filtering and refinement algorithms based on spatio-temporal consistency inherent in license plate image sequences. The experimental results in Subsection 4.6.2 demonstrates the effectiveness of the proposed filtering and refinement algorithms. Especially, Table 5 presents the results of ablation studies showing that our algorithms increase the character recognition accuracy by 11.74%p from 74.70% to 86.44%.

Leveraging the precisely aligned neighboring frames, MF-LPR² effectively restores severely degraded license plate images, as shown in Fig. 1. The most significant advantage of MF-LPR² over existing generative models is its ability to preserve the content of the input image by avoiding the generation of spurious artifacts and distortions. This feature is

essential for maintaining the evidential value of the original dash cam footage. Therefore, we propose a novel metric, the top-k Percentile average Distance to Nearest Frame (PDNF-k), to quantify the severity of local artifacts in the restored image.

In addition, we constructed the Realistic License Plate Restoration and Recognition (RLPR) dataset to evaluate the proposed framework. The RLPR dataset contains 200 low-quality road image sequences, each which consists of 31 consecutive frames captured by real dash cams. As a result, the RLPR dataset better reflects the image distortions found in real-world road scenarios compared to previous datasets, which consisted of high-resolution images and their artificially degraded versions. The RLPR dataset also includes pseudo ground-truth (GT) license plate images and text labels for each low-quality image sequence. The pseudo-GT image was created by extracting the license plate region from a high-quality frame within the same video clip as the low-quality image sequence, and then manually aligning it with the center frame among the 31 low-quality frames.

The main contributions of our work are summarized as follows:

- We report the inefficacy of conventional image restoration methods for license plate image restoration and address this limitation by proposing the MF-LPR² framework, which enhances legibility while preserving evidential content by leveraging multiple frames.
- Optical flow filtering and refinement algorithms that precisely align multiple low-quality image frames by leveraging spatio-temporal consistency.
- The realistic LPR dataset for evaluating multi-frame image restoration models under realistic conditions.
- A novel metric, the Top-k Percentile average of Distance to the Nearest Frame (PDNF-k) that quantifies the severity of spurious artifacts in the restored image.
- Significant improvements in image quality and recognition accuracy compared to 11 baseline methods.

2. Related Work

2.1. Image Restoration

License plate image restoration aims to enhance low-quality license plate images to a level where accurate recognition is possible. Previous work on license plate image restoration primarily addresses the motion blur problem. In contrast, in the field of general image restoration, various methods, such as SISR, MFSR, and VSR, have been developed to enhance the resolution and visual quality of poor quality images. The majority of these methods are based on conditional generative models and restore missing information in input images using knowledge obtained from training data. In this section, we introduce restoration methods applicable to license plate images and describe their limitations in license plate image restoration.

Single Image Super-Resolution

In recent years, the field of SISR has seen significant progress, evolving from CNN-based architectures to Transformer-based approaches. SRCNN (1) was one of the first deep-learning-based SR methods, utilizing a shallow CNN for high-resolution reconstruction. EDSR (4) improved upon this by removing batch normalization and increasing network depth. RCAN (5) introduced a Residual in Residual (RIR) structure and a channel attention mechanism, enabling deeper networks to focus on high-frequency information. More recent models, such as DCLS (14), leverage deformable convolutions to adaptively enhance different regions of the image.

SRGAN (2) improved perceptual quality by incorporating adversarial and perceptual loss functions. Subsequent models, including ESRGAN (3), Real-ESRGAN (7), and BSRGAN (8), further refined these methods by improving training stability and robustness to noise. Meanwhile, Transformer-based models have gained prominence: IPT (9) applied self-attention mechanisms to SR tasks, SwinIR (10) introduced Swin Transformers for better feature extraction, and Restormer (11) used channel-wise attention to boost efficiency. More recent methods, such as HAT (6) and DRCT (13), refine hierarchical attention strategies for high-fidelity image reconstruction.

Despite these notable advancements, existing SISR methods struggle to handle severe degradations and real-world noise characteristic of dash cam footage. Additionally, they are often trained on artificially degraded datasets, limiting their ability to generalize. To overcome these issues, our approach employs multi-frame information to surpass the inherent constraints of single-frame methods and validated on more realistic dataset.

In addition to general super-resolution techniques, several studies have explored SR approaches specifically for text images. TPGSR (23) introduced a text-prior-guided approach to enhance character restoration. STT (24) proposed a Transformer-based encoder for extracting text-specific features, while TATT (25) and DPMN (26) introduced further refinements in text-specific feature extraction and reconstruction. However, these methods primarily address generic scene text and do not fully account for the unique challenges of license plates, such as severe motion blur, reflection light, and poor resolution. Our method specifically addresses license plate restoration by incorporating multi-frame alignment to handle severe degradation that is difficult to restore from a single image.

Multi-frame Super Resolution

MFSR models enhance image quality by aligning and aggregating information across consecutive frames. In recent models, optical flow estimators are commonly employed for frame alignment. Early optical flow-based techniques, such as the Lucas-Kanade algorithm (27), relied on handcrafted feature extractors. With the advent of deep-learning, CNN-based models like FlowNet (28) and SpyNet (29) demonstrated improved accuracy in flow estimation. More recently, Transformer-based approaches such as FlowFormer (30) extended SpyNet by integrating self-attention mechanisms, and FlowFormer++ (31) further refined this architecture to boost flow estimation performance. While these approaches generally excel in multi-frame alignment, they often struggle under extreme low-quality conditions, leading to alignment errors and degraded final image quality. Building on optical flow-based alignment, our method introduces a novel spatio-temporal filtering and refinement mechanism for multi-frame aggregation, ensuring more accurate alignment and superior restoration results.

Video Super Resolution

VSR aims to reconstruct high-quality frames by leveraging information from multiple consecutive frames. VSR is distinguished from MFSR in that it restores multiple frames, whereas MFSR focuses on restoring a single image. EDVR (17) introduced deformable convolutions to enhance feature aggregation, while BasicVSR++ (15) advanced recurrent networks through improved temporal feature propagation. Later approaches, such as RVRT (19), employed Transformer-based recurrent learning to refine temporal dependencies, and IART (20) further enhanced temporal consistency by reducing misalignment errors. Although these methods exploit multi-frame information, they often rely on synthetic datasets where degradation is artificially induced, limiting their performance in real-world scenarios. Our approach addresses this challenge by integrating a spatio-temporal filtering and refinement mechanism that mitigates alignment errors and boosts robustness against real-world noise, particularly in dash cam footage.

2.2. License Plate Image Restoration and Recognition

In this section, we introduce previous work on license plate recognition and image restoration techniques specially designed for license plate recognition.

License Plate Recognition

Lic ense plate recognition (LPR) involves both detection and character recognition. Numerous detection approaches have been proposed, ranging from general object detection frameworks such as Faster R-CNN (32), SSD (33), and YOLO (34), to salient detection models like APNet Salient (37), WaveNet Salient (38), LSNet (39) which focus on identifying the most critical objects. Among these, LPR research has progressed in tandem with the development of general object detection methods. For instance, WPODNet (40), built upon YOLO, was proposed for LPR but often struggles to deliver a reliable performance under real-world conditions involving low-quality license plate images.

License Plate Image Restoration

Previous work on license plate image restoration have mainly tackled motion deblurring (41; 42; 46; 45). Early studies on motion deblurring primarily estimated blur kernels using convolutional neural network (CNN)-based methods (42). Sun et al. (41) introduced a deep-learning framework explicitly designed to model non-uniform motion blur distributions at the patch level, setting a foundation for CNN-based motion deblurring techniques. Later, generative models such as Kupyn et al. (46) and Nguyen et al. (45) employed generative adversarial networks (GAN) to remove motion blur while preserving realistic textures. However, these models often struggle with recovering fine details in high-frequency areas such as small texts on license plates, leading to excessive smoothing and loss of crucial information.

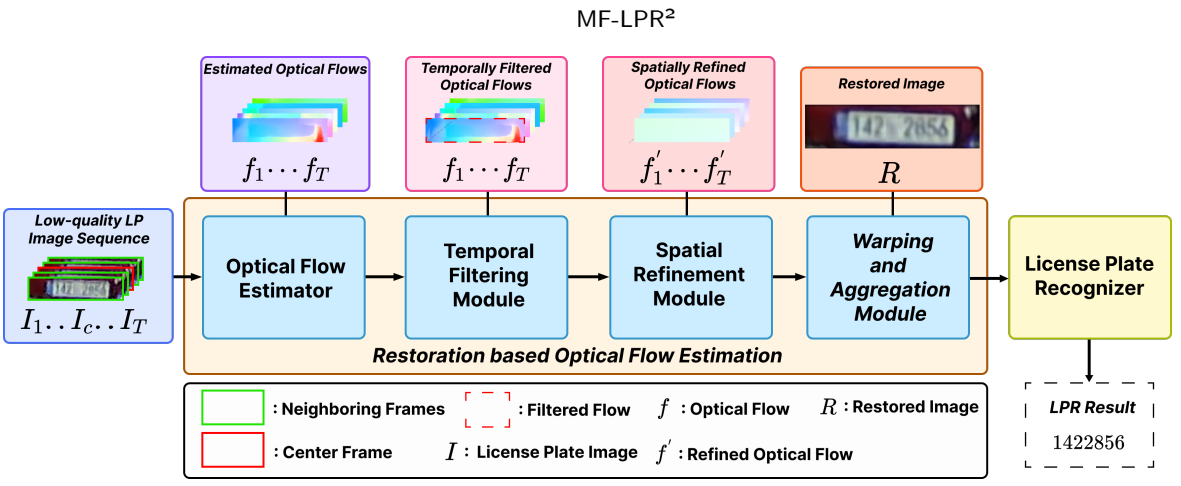

Figure 2: The overall structure of the MF-LPR² framework.

Later, Zou et al. (47) restored license plate images of extremely low-resolution by leveraging textual priors, similar to STISR methods. Nascimento et al. (48) proposed a Transformer-based SISR method for low-resolution license plates, utilizing attention modules and an OCR-guided perceptual loss to enhance recognition accuracy. A more recent approach, AFA-Net (49), integrated a super-resolution network and a deblurring network to achieve better restoration results. Nevertheless, these single-frame approaches do not take advantage of temporal information, which is critical for addressing severe degradation in real driving environments. Moreover, these methods were focused on artificially degraded datasets, limiting their ability to generalize to real-world noise. Our method extends these approaches by utilizing multi-frame information to achieve superior restoration performance.

The study most closely related to our work among previous research is 'Eyes on the Target' (50). However, it was developed long ago and relies on a traditional flow estimation algorithm to align frames, which limits its effectiveness in complex real-world scenarios. In contrast, our method exhibits significantly higher performance by incorporating a state-of-the-art neural flow estimator and novel flow filtering refinement methods. Table 3 presents a comparative analysis demonstrating that MF-LPR² outperforms Eyes on the Target by large margins in terms of both image quality and recognition rate.

3. Method

3.1. Overview

The proposed MF-LPR² restores and recognizes a high-quality license plate image from a sequence of low-quality image frames captured by consumer-grade cameras, such as dash cams. MF-LPR² consists of a restoration module, and a license plate recognizer, as illustrated in Fig. 2. Firstly, the restoration module restores the resolution and quality of the temporal center frame by combining the complementary information from its neighboring frames. Then, the license plate recognizer predicts the text from the restored image.

3.2. Restoration Module

The restoration module is a core component for recognizing low-quality license plate images. It enhances the temporal center frame, I_c, by aligning the neighboring frames (I_t's with $t \neq c$) to I_c and then aggregating the aligned frames to produce a high-quality image. For alignment, MF-LPR² estimates optical flow between the center frame and each neighboring frame, then warps the neighboring frames to match the center frame using the estimated optical flow. To precisely estimated optical flow, we combine a state-of-the-art optical flow estimator and novel filtering and refinement algorithms. The filtering algorithm rejects optical flows with severe errors based on temporal consistency, while the refinement algorithm detects and corrects local errors in the estimated optical flow based on spatial consistency.

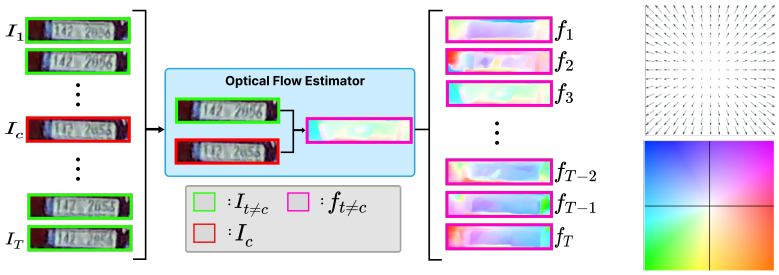

Figure 3: Optical Flow Estimation: Given a sequence of license plate images, MF-LPR² estimates the optical flow from the center frame to each neighboring frame. The direction represented by the colors of the optical flow corresponds to the color coding chart on the right.

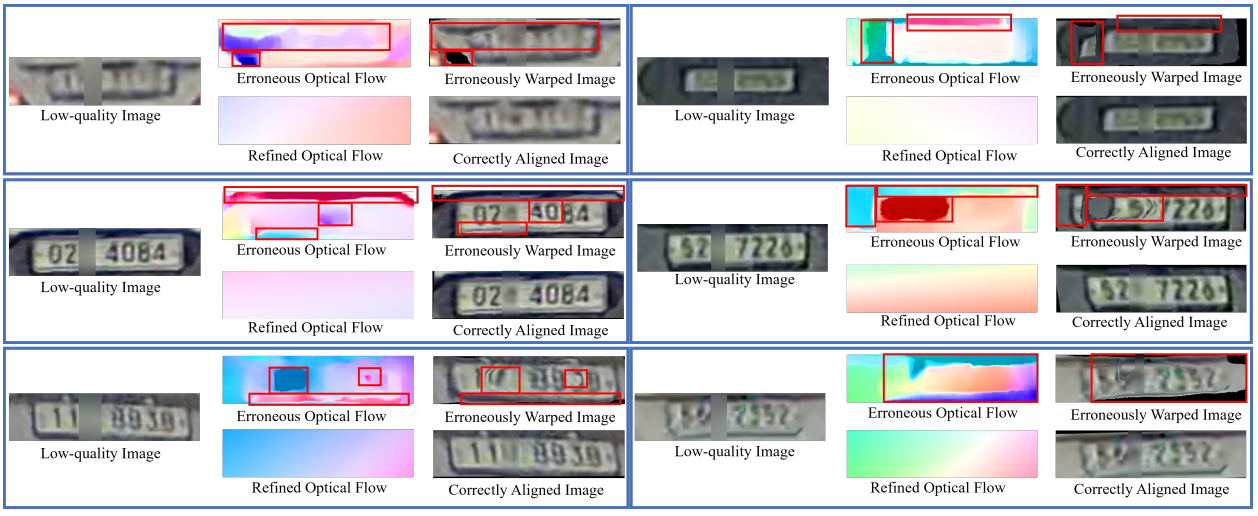

Figure 4: Comparison of erroneous and refined optical flows and their corresponding alignment results. Each blue box displays a low-quality image (left), the estimated and refined optical flows (center), and the alignment results based on the optical flows (right). The red boxes highlight errors. Even a state-of-the-art estimator, FlowFormer++ (31), produces substantial errors, which severely degrade the alignment results. However, the proposed method successfully refines the optical flows, significantly reducing the alignment errors as a result.

3.2.1. Optical Flow Estimator

In a dash cam video clip, the location and orientation of the license plate change over time. An optical flow represents the displacement of each pixel between a pair of frames. The estimator takes the center frame I_c and a neighboring frame I_t as input, and estimates optical flow f_t from I_c to each I_t, as Eq. 1, where $H_c \times W_c$ is the size of I_c, $u_t \in \mathbb{R}^{H_c \times W_c}$ and $v_t \in \mathbb{R}^{H_c \times W_c}$ are the vertical and horizontal displacements from I_c to I_t, respectively, i.e., $I_c(i,j) \approx I_t(i + u_t(i,j), j + v_t(i,j))$.

$$f_t = (u_t, v_t) = \text{OpticalFlow}(I_t, I_c) \tag{1}$$

Recently, Huang et al. (30) proposed a Transformer-based optical flow estimator, FlowFormer, which tokenizes a 4D cost volume from an input image pair, encodes the cost volume into a cost memory, and estimates optical flow by decoding the cost memory using a recurrent Transformer decoder. More recently, Shi et al. (31) enhanced FlowFormer by pretraining the cost-volume encoder with a masked cost volume autoencoder, resulting in FlowFormer++. In this study, we apply the FlowFormer++ to estimate optical flows between the center and neighbor frames. Although FlowFormer++ is a state-of-the-art optical flow estimator that demonstrates high-performance on high-quality images, it often produces erroneous results on low-quality license plate images, as shown in Fig. 4.

3.2.2. Temporal Filtering Module

The temporal filtering module rejects the optical flows that are poorly estimated overall based on temporal consistency between adjacent optical flows. Since the interval between frames is short, the differences between adjacent optical flows are moderate. Therefore, we reject an optical flow f_t if it differs significantly from the adjacent optical flows, i.e., Average($|f_{t+1} - f_t|$) > θ_{temp} and Average($|f_t - f_{t-1}|$) > θ_{temp}, where θ_{temp} is a threshold value. In this study, we set $\theta_{temp} = 10$ according to the experimental result presented in Fig. 11. For the first and the last frames ($t = 1$ and $t = T$), we compare f_t with only one adjacent optical flow. When an optical flow f_t is rejected, we discard the corresponding frame I_t. It is noteworthy that the temporal filtering module only rejects completely misestimated optical flows, and does not reject those that are partially misestimated, as rejections are based on the average difference.

We have attempted to refine misestimated optical flows rather than rejecting them. For example, we replaced misestimated optical flows f_t's by the interpolation of the previous and next optical flows, $(f_{t-1} + f_{t+1})/2$. However, such refinement was ineffective in improving output quality because the frames with overall misestimated optical flows were usually too poor in quality to be used for restoring the center frame.

3.2.3. Spatial Refinement Module

Although the temporal filtering module rejects optical flows that are highly inconsistent with adjacent frames, the remaining optical flows still contain errors. Instead of aligning neighboring frames directly using the estimated optical flows, we refine them based on spatial consistency. Since a license plate is a rigid object with a planar shape, optical flows between license plate images are spatially smooth, this means that the values in u_t and v_t change linearly with the coordinates. Leveraging this spatial consistency, we refine the estimated optical flows to be planar through linear approximation, and use the resulting planar optical flow to align the frames.

To this end, we fit the estimated optical flow to a planar model. An important requirement for the fitting algorithm is that it should be robust to local errors in the estimated optical flow. In order to prevent local errors from affecting the refinement process, we fit the estimated optical flow to a plane based on the horizontal and vertical median values, which are tolerant to outlier values. First, we find the horizontal median values for each row and the vertical median values for each column, as Eq. 2. As the spatial refinement is performed for each frame, we omit the time index t for simplicity in this section.

$$\begin{aligned} Med_{hor}(f) &= (Med_{hor}(u), Med_{hor}(v)) \\ Med_{ver}(f) &= (Med_{ver}(u), Med_{ver}(v)) \end{aligned} \tag{2}$$

$Med_{hor}(u) \in \mathbb{R}^{H' \times 1}$ and $Med_{hor}(v) \in \mathbb{R}^{H' \times 1}$ are the horizontal median values of u and v for each row, while $Med_{ver}(u) \in \mathbb{R}^{1 \times W'}$ and $Med_{ver}(v) \in \mathbb{R}^{1 \times W'}$ are the vertical median values of u and v for each column. To further reduce the influence of outliers, we exclude the top and bottom 15% of values and then find the median of the remaining values.

With the horizontal and vertical median values, we fit $Med_{hor}(u)$, $Med_{ver}(u)$, $Med_{hor}(v)$, and $Med_{ver}(v)$ to lines as Eq. 3, where α_* and β_* are the coefficients for horizontal median values while γ_* and δ_* are those for vertical median values. The coefficients are estimated from the median values by linear approximation.

$$\begin{aligned} R_u(i,j) &= (R_{hor}(u), R_{ver}(u)) = (\alpha_u i + \beta_u, \gamma_u j + \delta_u) \\ R_v(i,j) &= (R_{hor}(v), R_{ver}(v)) = (\alpha_v i + \beta_v, \gamma_v j + \delta_v) \end{aligned} \tag{3}$$

Then, we compute the refined optical flow $f' = (u', v')$ using Eq. 3, as Eq. 4.

$$\begin{aligned} f'(i,j) &= (u'(i,j), v'(i,j)) \\ u'(i,j) &= \alpha_u(i - H/2) + \gamma_u j + \delta_u \\ v'(i,j) &= \alpha_v(i - H/2) + \gamma_v j + \delta_v \end{aligned} \tag{4}$$

Fig. 4 presents examples of estimated and refined optical flows along with the corresponding alignment results. Each blue box displays a low-quality image (left), estimated and refined optical flows (center), and the alignment results (right). While the estimated optical flows contain substantial errors, producing erroneous alignment results, the proposed method demonstrates significantly improved results.

However, the refined optical flow f' does not perfectly approximate the real optical flow because it is based on the assumption that the optical flow between license plate images is planar. Moreover, it estimates the coefficients from

the median values of the estimated optical flow, which may contain error. Therefore, the difference between f and the planar model f', $\mathrm{Diff}(f, f')$, reflects not only estimation error in f but also approximation error in f', making the substitution of f_t with f'_t for all t sub-optimal.

Therefore, we detect erroneous optical flows by comparing the maximum and median of $\mathrm{Diff}(f_t, f'_t)$ as Eq. 5, where the maximum value represents the error in high-error areas, while the median value represents the error in other areas. The former reflects estimation error, whereas the latter reflects approximation error. We set $\theta_{spatial} = 20$ according to the experimental result presented in Fig 12.

$$\max_{i,j}(\mathrm{Diff}(f(i,j), f'(i,j))) - \mathrm{median}_{i,j}(\mathrm{Diff}(f(i,j), f'(i,j))) > \theta_{spatial} \tag{5}$$

3.2.4. Warping and Aggregation Module

With the refined optical flows, we align and aggregate the neighboring frames to produce a high-quality image. For this, we apply the Geometric k-nearest neighbors Super-Resolution (GSR_4) algorithm (56). First, it places the pixels from all frames I_t onto the coordinate plane of the output image, which is the same as that of the center frame in MF-LPR² except that it allows fractional coordinates. It transforms the coordinate of each pixel in I_t to the corresponding output coordinate using the optical flow. Then, for each integer coordinate (i, j) on the output coordinate plane, the it finds the nearest pixels in each frame. Finally, it fills each output pixel $\hat{Y}(i, j)$ with the average of the nearest neighbor pixels for each (i, j).

It is noteworthy that, since GSR_4 determines pixel values based on the average of aligned neighboring frames, it effectively suppresses noise and outlier pixels without introducing structural patterns that deviate significantly from those of the input images. As a result, as shown in Fig. 6, MF-LPR² does not introduce severe artifacts or distortions, thereby avoiding arbitrary alterations to the evidential contents of the input images. In license plate recognition, this represents a significant advantage compared to generative models that actively utilize pretrained priors to restore images.

3.3. License Plate Recognizer

To segment and recognize the characters in the restored license plate image, we utilize an off-the-shelf scene text recognizer, MGP-STR (57), which is both efficient and effective in handling low-quality images. MGP-STR demonstrates excellent performance in managing diverse text appearances, varying orientations, and complex backgrounds, making it particularly well-suited for recognizing characters in license plate images captured in real-world scenarios.

3.4. The Realistic License Plate Restoration and Recognition Dataset

Since MF-LPR² takes a sequence of road image frames as input, its evaluation requires a dataset of low-quality road image sequences that resemble real-world driving conditions. Additionally, for quantitative evaluation, a high-quality pseudo ground truth (GT) image and text label for each image sequence are essential. However, existing license plate datasets do not satisfy these specific requirements.

Limitations of Existing Datasets

Several existing license plate datasets, such as OpenALPR-EU (60), SSIG-SegPlate (61), UFPR-ALPR (62), PKU-SR (63), and RodoSol-SR (63), have been introduced primarily for Automatic License Plate Recognition (ALPR). While these datasets feature images from diverse environments, they are not well-suited for evaluating multi-frame restoration algorithms for low-quality license plate images. OpenALPR-EU focuses on single-frame images of parked vehicles, which fail to represent real-world driving conditions. SSIG-SegPlate and UFPR-ALPR include multi-frame sequences but generally comprise high-quality images, not reflecting the challenges of real-world low-quality data. Datasets like PKU-SR and RodoSol-SR rely on artificial degradation methods, such as Gaussian noise and bicubic downsampling, to generate low-quality frames.

Fig. 5 (a) illustrates the conventional process of generating high- and low-quality image pairs through artificial degradation. This process applies multiple degradation steps, including noise addition, downsampling, and lossy compression. However, artificially degraded images often fail to mimic the diverse and complex degradations found in real-world conditions. Furthermore, the consistent level of degradation across all frames reduces inter-frame complementarity, making these datasets less effective for evaluating multi-frame restoration algorithms.

Dataset Construction Process

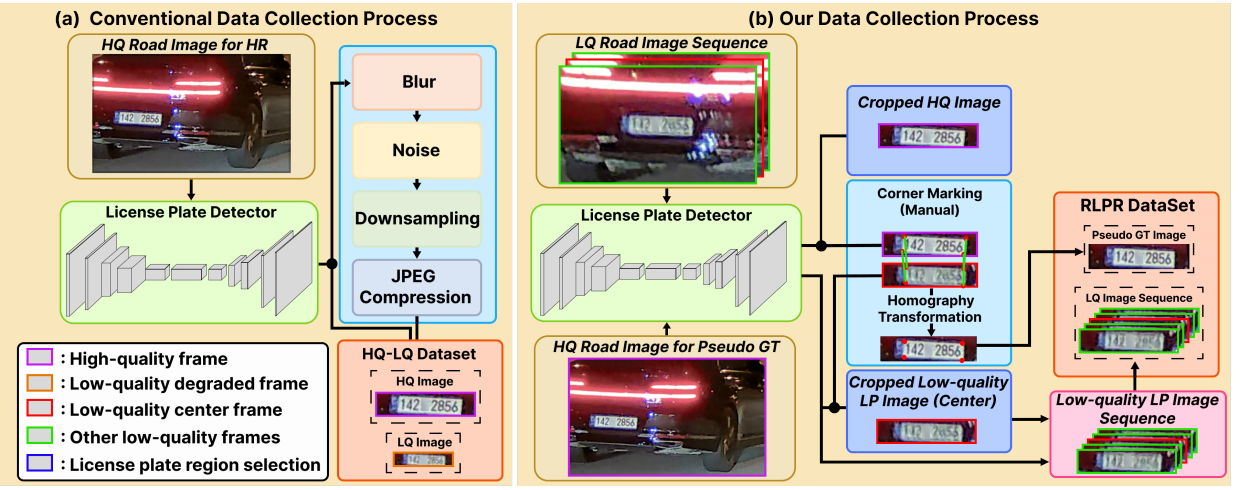

Figure 5: The process of collecting high- and low-quality image pairs: the conventional process based on artificial degradation (Left) and the process used to collect the RLPR dataset (Right). The RLPR dataset consists of real low-quality images that reflect the complexities of real-world scenarios.

The building process of the RLPR dataset (as illustrated in Fig. 5 (b)) ensures that the dataset closely mirrors real-world driving conditions while maintaining high-quality annotations. From 1,052 dash cam video clips, we extracted sequences with visible license plates exhibiting sufficiently low quality. We also extracted a high-quality frame for the pseudo-GT image from the same video clip. Only 200 sequences were retained after meticulous manual screening to ensure they included both low-quality frames and high-quality frames. For each sequence, the following steps were performed:

First, a state-of-the-art DeepLabV3 (53) model with a ResNet-101 backbone, fine-tuned on 130 manually annotated road images, was used to detect license plate regions. These regions were manually refined and cropped with slightly expanded bounding boxes to preserve important details, resulting in slight positional and size variations across frames that reflect real-world conditions.

Pseudo-GT images were then selected as the highest-quality frames from the same video clips but were often temporally distant from the low-quality sequences. To ensure precise alignment, the corners of the license plate regions were manually marked on both the pseudo-GT and the center frame of the sequence, followed by a homography transformation. The aligned GT images retained only the license plate region, with background areas removed, ensuring accurate comparability with the low-quality frames.

This rigorous construction process ensures that RLPR captures the complexities of real-world conditions, overcoming the limitations of artificially degraded datasets. Unlike other datasets that rely on artificial degradation, RLPR provides sequences with diverse, dynamic degradations and complementary frame information, making it uniquely suited for evaluating multi-frame restoration frameworks.

The RLPR dataset encompasses a diverse range of conditions, including lighting (daytime and nighttime), lane positions (ego and non-ego lanes), and license plate types (painted and reflective), as shown in Table 1. This diversity ensures that RLPR serves as a robust benchmark for realistic performance evaluation. Furthermore, by pairing each sequence with carefully aligned high-quality pseudo-GT images, RLPR enables researchers to perform precise quantitative analysis of restoration and recognition tasks. Consequently, despite the limited number of samples, which makes the RLPR dataset not suitable for model training, it possesses high value as an evaluation dataset.

3.5. Top-k Percentile Average Distance to Nearest Frame (PDNF-k)

A critical and unique requirement in license plate image restoration is that the evidential contents of the original image must not be compromised. However, many restoration models, which actively utilize priors obtained from training data, frequently generate spurious artifacts, as shown in Fig. 6. The artifacts significantly degrade the essential information required to preserve character class features.

Nevertheless, since artifacts occupy only a small portion of the overall image, conventional quality assessment metrics that reflect the global difference with the GT image fail to adequately capture the impact of these artifacts.

Table 1
Composition of RLPR dataset. Each sample is categorized by lighting condition (daytime/nighttime), lane position (ego/non-ego), and plate type (painted/reflective film).

Category	Number of Samples	Number of Characters
RLPR Dataset	200	1261
Daytime	119	715
Nighttime	81	510
Ego Lane	42	263
Non-ego Lane	158	998
Painted	165	1016
Reflective Film	35	245

Common metrics like PSNR, SSIM, and LPIPS integrate local differences via spatial averaging or Minkowski pooling. PSNR is derived from the mean squared error (MSE) between the GT and output images. SSIM measures structural similarity through Minkowski pooling across spatial and frequency domains. LPIPS evaluates perceptual similarity by averaging embedding differences across spatial dimensions and layers. While these metrics capture overall differences, they are not sensitive to localized spurious artifacts.

To complement such a limitation of conventional metrics, we propose a novel metric to quantify the degradation of evidential contents caused by spurious artifacts. An effective metric that quantifies the severity of spurious artifacts must meet the following conditions:

1. To measure the degradation of original information rather than restoration performance, it should be computed based on the difference from the input image, rather than the GT image.
2. Since low-quality input images reflect pixel values with uncertainty, the metric should account for this by representing each pixel value as a distribution. This helps mitigate unnecessary penalties that correctly restored images may otherwise receive.
3. It should focus on regions with significant discrepancies rather than simply integrating local differences across spatial dimensions.

Regarding conditions 1 and 2, we estimate the distribution of each input pixel (i, j) based on the corresponding pixel values from multiple frames. However, modeling the distribution of pixel values as a Gaussian distribution is inappropriate for two reasons. First, when the aligned frames are similar to each other, the variation is measured to be excessively low, leading to an underestimation of pixel uncertainty. Second, since GSR_4 applied to MF-LPR² outputs the mean value, a Gaussian model would overly favor MF-LPR². Therefore, we adopt a non-parametric approach similar to the k-nearest neighbor classifier. Specifically, the error for each output pixel (i, j) is defined as the distance to the nearest corresponding pixels in the aligned input frames.

$$\mathrm{Dist}_t(i,j) = \left|\hat{Y}(i,j) - I'_t(i,j)\right| \tag{6}$$

$$\mathrm{DNF}(i,j) = \min_t \mathrm{Dist}_t(i,j), \tag{7}$$

where I'_t is the t-th aligned input frame as $I'_t(i,j) = I_t(i + u_t(i,j), j + v_t(i,j))$.

For condition 3, we integrate pixel-wise errors $\mathrm{DNF}(i,j)$ using the top-k percentile average as

$$\text{PDNF-}k = \frac{1}{|\Omega_k|} \sum_{(i,j)\in\Omega_k} \mathrm{DNF}(i,j), \tag{8}$$

where Ω_k is the set of pixels with the highest $k\%$ DNF values.

The key characteristics of PDNF-k are as follows:

- A large PDNF-k value indicates that the output image contains severe spurious artifacts.
- A smaller PDNF-k value suggests that the information in the original image is better preserved. However, an excessively small PDNF-k value may imply a limited enhancement effect on output quality. For instance, if the model produces an output identical to the low-quality center frame, PDNF-k is measured as zero.

Table 2
Image quality evaluation results of the proposed model, MF-LPR², compared with other recently developed baseline models. ↑ indicates higher values are better, while ↓ indicates lower values are better. The best and second-best performances are highlighted in bold and underlined, respectively. MF-LPR² outperformed the baseline models in all metrics.

Method	Single/Multi Frame	PSNR↑	SSIM↑	LPIPS↓	PDNF-5↓	LPR Accuracy↑
SwinIR (10)	Single Frame	14.6156	0.2931	0.5495	108.12	3.49%
DCLS (14)	Single Frame	15.6955	0.3121	0.5908	27.60	16.18%
HAT (6)	Single Frame	15.6339	0.3064	0.6021	35.50	15.07%
DRCT (13)	Single Frame	15.6031	0.3023	0.5942	37.94	15.07%
TATT (25)	Single Frame	14.8347	0.3149	0.5393	69.80	9.36%
DPMN (26)	Single Frame	15.6304	<u>0.3383</u>	<u>0.5036</u>	40.11	14.99%
RVRT (19)	Multi Frame	15.7743	0.3155	0.5595	**6.96**	18.56%
IART (20)	Multi Frame	<u>15.7951</u>	0.3149	0.5565	<u>9.58</u>	<u>18.95%</u>
MF-LPR²	Multi Frame	**16.3478**	**0.3486**	**0.4629**	14.00	**86.44%**

- PDNF-k is a metric designed to complement existing quality assessment metrics and is intended to be used alongside other metrics.
- As k decreases, PDNF-k focuses more on pixels with large errors, whereas as k increases, it reflects overall error.

4. Experiments

4.1. Experimental Settings and Evaluation Metrics

Our framework consists of two main components: a restoration module and a license plate recognizer. For the optical flow estimator in the restoration module and for the license plate recognizer, we used the official implementations of FlowFormer++ (55) and MGP-STR (58), respectively. All experiments were conducted on a server equipped with an NVIDIA RTX 2080.

For evaluation, we first restored and recognized license plate areas in the RLPR dataset by MF-LPR². We then evaluated the image restoration performance by comparing the restored images with the pseudo-GT images. For quantitative evaluation, we assessed the quality of the restored images using metrics such as Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index (SSIM) (51), and Learned Perceptual Image Patch Similarity (LPIPS) (52). These metrics were computed only for the pixels within the license plate region, excluding the background areas. Additionally, we measured the severity of spurious artifacts in PDNF-5 introduced in Section 3.5. Finally, we evaluated the recognition performance by measuring the character recognition accuracy using the text labels from the RLPR dataset. For comparison analysis, we compared MF-LPR² with eight recently-developed image/video restoration models, (SwinIR, DCLS, HAT, DRCT, DPMN, TATT, RVRT, and IART) and three license plate recognition (LPR) models (WPOD-Net, AFA-Net, and Eyes on the Target). The performance of the baseline models was measured using their open-source implementations and pretrained parameters.

4.2. Comparison with Image/Video Super-Resolution Models

We compared the performance of MF-LPR² against eight state-of-the-art image and video super-resolution models: SwinIR(10), DCLS(14), HAT(6), DRCT(13), DPMN(26), TATT(25), RVRT(19), and IART(20). The evaluation was carried out using PSNR, SSIM, LPIPS, PDNF-5 and character recognition accuracy metrics. For character recognition accuracy, we integrated each restoration model with the same recognizer used for MF-LPR².

Table 2 presents the evaluation results. The poor quality of input images caused the baseline models to produce unreliable results, often leading to spurious artifacts and low recognition accuracy. Fig. 6 illustrates the restoration results of several samples. MF-LPR² showed significantly higher performance in all metrics, achieving a PSNR of 16.3478, a SSIM of 0.3486, and a LPIPS of 0.4629. Notably, MF-LPR² demonstrated a character recognition accuracy of 86.44% which is remarkably higher than the second best performance of 18.95% achieved by IART. These results show the effectiveness of MF-LPR² in restoration and recognition of poor quality license plate images.

Regarding PDNF-5, MF-LPR² showed the third lowest value, following RVRT and IART. As shown in Fig. 6, these three models do not generate spurious artifacts significantly. However, RVRT and IART produced output images that

Table 3
Evaluation results of MF-LPR² compared with a license plate image restoration and recognition framework, WPOD-Net, AFA-Net, and Eyes on the Target. ↑ indicates higher values are better, while ↓ indicates lower values are better. MF-LPR² outperformed the baseline models in all metrics. As WPOD-Net model is a recognition model without a restoration module, its SSIM, PSNR, and LPIPS calculation are omitted. The high PDNF-5 value of Eyes on the Target is due to the official code, which generates rectified images that may be misaligned relative to the input image.

Method	Single/Multi Frame	PSNR↑	SSIM↑	LPIPS↓	PDNF-5↓	LPR Accuracy↑
WPOD-Net (40)	Single Frame	-	-	-	-	13.08%
AFA-Net (49)	Single Frame	14.7506	0.3018	0.6110	87.85	14.04%
Eyes on the Target (50)	Multi Frame	14.9170	0.2680	0.5126	117.77	82.55%
MF-LPR²	Multi Frame	**16.3478**	**0.3486**	**0.4629**	**14.00**	**86.44%**

overly resemble the input images. As a result, they performed worse than MF-LPR² in other metrics. SwinIR, TATT, and DPMN exhibited significantly high PDNF-5 values indicating that they suffer severely from artifacts, as shown in Fig. 6.

4.3. Comparison with License Plate Image Restoration and Recognition Models

We also compared MF-LPR² with three license plate image restoration and recognition models presented in previous studies, WPOD-Net(40), AFA-Net(49), and Eyes on the Target(50), which were specifically designed for license plate images. Table 3 presents the comparison results. The single-frame LPR frameworks, WPOD-Net and AFA-Net, achieved an LPR accuracy of 13.08%, showing no significant improvement compared to MGP-STR combined with the restoration models, shown in Table 2.

However, the multi-frame LPR model, Eyes on the Target, achieved significantly higher character recognition accuracy than the single-frame LPR frameworks. MF-LPR² exhibited even higher performance than Eyes on the Target across all metrics largely due to its accurate estimation of optical flows via the proposed filtering and refinement algorithms. MF-LPR² also outperformed all baseline LPR models in PSNR, SSIM, and LPIPS by large margins.

MF-LPR² also exhibited a lower PDNF-5 value than the other models. Eyes on the Target showed a high PDNF-5 value, while its results in Fig. 6 contain little artifacts. The high PDNF-5 value is attributed to the official code, which generates rectified images that may be misaligned relative to the input image. This may also affect the PSNR value, whereas SSIM, LPIPS, and LPR accuracy are less sensitive to slight misalignment. For the reliability of the experiment, we did not modify the official code. The PDNF-k values of the models for $k = 5, 10, 15, 20$ are compared in Fig. 7. The graph exhibits a similar trend for all k. However, the differences between models are most pronounced in PDNF-5, which focuses the most on severe artifacts.

4.4. Qualitative Evaluation

In addition to the quantitative evaluation using the pseudo-GT images, we qualitatively compared the restoration results of MF-LPR² with those of the eleven baseline models, SwinIR, DCLS, HAT, DRCT, DPMN, TATT, RVRT, IART, WPOD-Net, AFA-Net, and Eyes on the Target. Fig. 6 displays the restoration results of MF-LPR² and the baseline restoration methods. When applied to low-quality images, the super-resolution models did not effectively enhance their visual quality. Especially, they often produced spurious artifacts as highlighted in the red boxes. For example, the character '6' in the second input image was severely degraded, making it easily confused with '0', '5', or '3.' The eight baseline methods (SwinIR, DCLS, HAT, DRCT, DPMN, TATT, RVRT, and IART) did not reduce ambiguity but restored it similarly to character '5' or '3.' The output quality of Eyes on the Target was the best among the baseline models, but it produced more blurry outputs compared to MF-LPR². In contrast, the proposed model, MF-LPR², yielded significantly superior results by reducing ambiguity and enhancing visual readability.

4.5. Model Complexity

We measured the complexity of the license plate restoration and recognition models. Table 4 summarizes the comparison results for these models. The number of parameters and the amount of compute in FLOPs were measured using the PyTorch Profiler. WPOD-Net, a single frame-based recognition model, is the lightest and fastest in terms of computational cost among single frame models; however, its performance is significantly lower than other models.

Figure 6: The restoration results of MF-LPR² compared with baseline models. The red boxes highlight artifacts and distortions due to the poor quality of the input image. We blurred the third letter in each image to de-identify personal information.

Table 4
Complexity Analysis between License Plate Image Restoration and Recognition Models. Inference Time is calculated per frame.

Method	Single/Multi Frame	Params (M)	FLOPs (G)	Inference Time (s)
WPOD-Net (40)	Single Frame	1.64	3.73	0.18
AFA-Net (49)	Single Frame	139.64	481.62	0.55
Eyes on the Target (50)	Multi Frame	-	-	3
MF-LPR²	Multi Frame	163.76	71.57	5.7

In contrast, AFA-Net incorporates both a super-resolution network and a deblurring network, resulting in the highest FLOPs among all models despite being single frame-based, yet its performance remains limited. For Eyes on the Target — a framework based on traditional algorithms — only the inference time was evaluated. Our proposed MF-LPR² model strikes a reasonable balance by achieving a relatively low FLOPs-to-parameter ratio. While its inference time

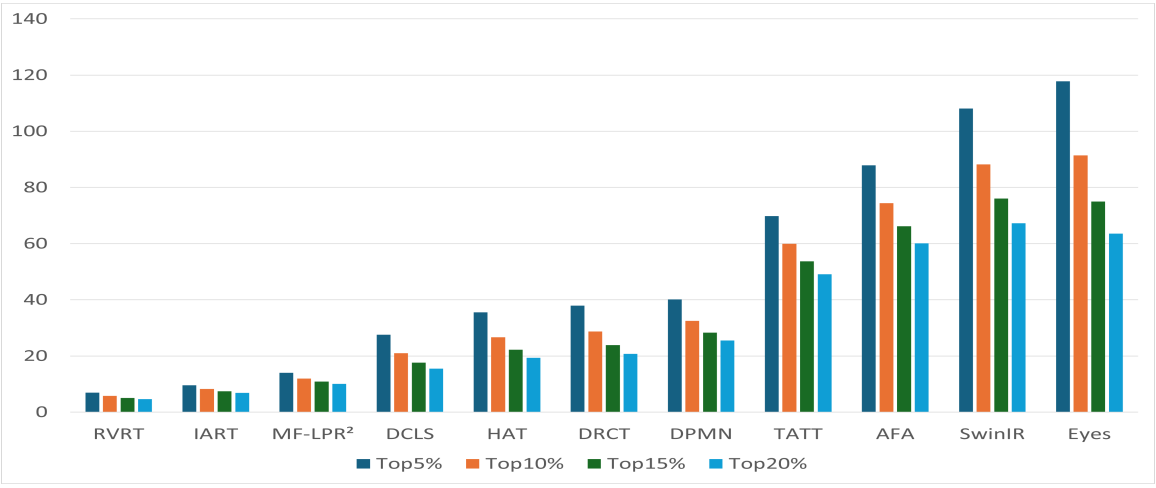

Figure 7: The PDNF-k of restoration models averaged over the RLPR dataset with $k = 5, 10, 15, 20$.

is somewhat higher compared to other models, it demonstrates superior overall performance, making it an effective solution for license plate restoration and recognition tasks.

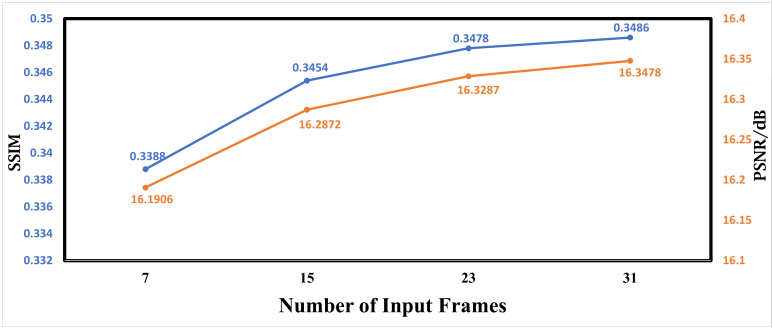

Figure 8: The performance of MF-LPR² with varying numbers of input frames.

4.6. Ablation Studies

We performed ablation studies to analyze the impact of the choices for each component on performance. We compared the performance of MF-LPR² by modifying various design options such as the number of frames in the input sequence, existence of modules, the threshold values for the temporal filtering and spatial refinement of the optical flow.

4.6.1. The Number of Frames in Input Image Sequence

We evaluated the restoration performance of MF-LPR² in SSIM and PSNR by changing the number of low-quality images in the input sequence. Fig. 8 displays the average SSIM and PSNR values of the output images restored from 7, 15, 23, and 31 input frames. Both SSIM and PSNR values consistently increase with the number of input frames. These results suggests that MF-LPR² effectively utilizes the information of multiple input frames to enhance the quality of the center frame.

4.6.2. Temporal Filtering and Spatial Refinement Modules

To better understand the contributions of the temporal filtering and spatial refinement modules to the overall performance of MF-LPR², we conducted a controlled experiment by developing two model variants and measuring their performance. The first variant only includes the optical flow estimation step and excludes both the temporal

Table 5
Performance comparison of MF-LPR² with different module configurations. **F**, **T**, and **S** represent optical flow estimation, temporal filtering, and spatial refinement, respectively. Both temporal filtering and spatial refinement improve PSNR, SSIM, and recognition accuracy.

F	T	S	PSNR↑	SSIM↑	LPR Accuracy↑
✓			16.3114	0.3468	74.70%
✓	✓		16.3240	0.3480	75.97%
✓	✓	✓	**16.3478**	**0.3486**	**86.44%**

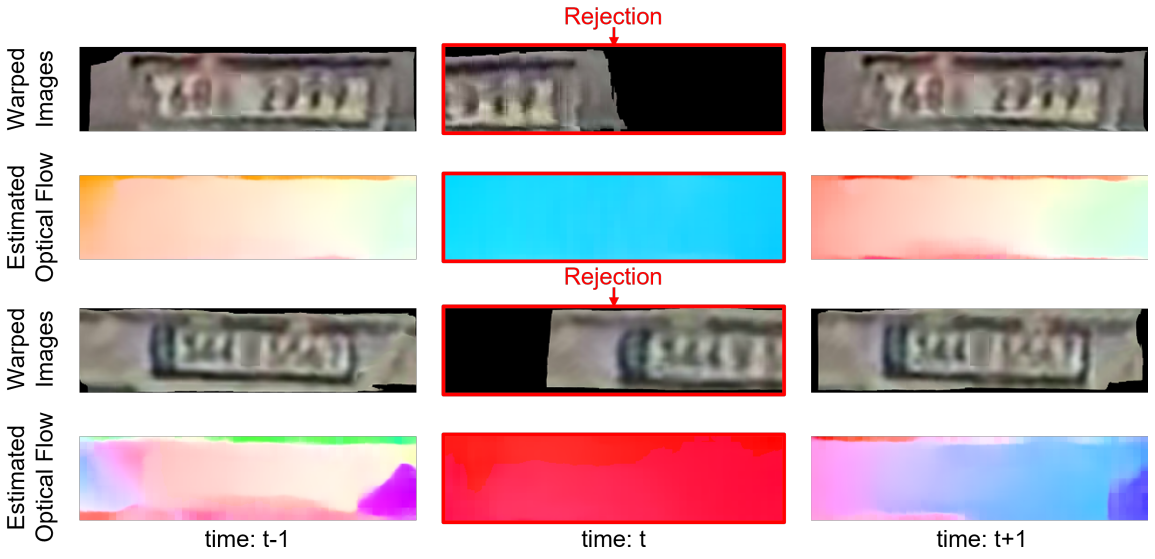

Figure 9: Qualitative results of the Temporal Filtering Module in MF-LPR². Our module detects and rejects completely misestimated optical flows, improving LPR accuracy by 1.27%p. This enhances the inter-frame consistency required for recognition tasks. The solid blue and solid red colors represent misestimated directions, with substantial biases to the left and right, respectively.

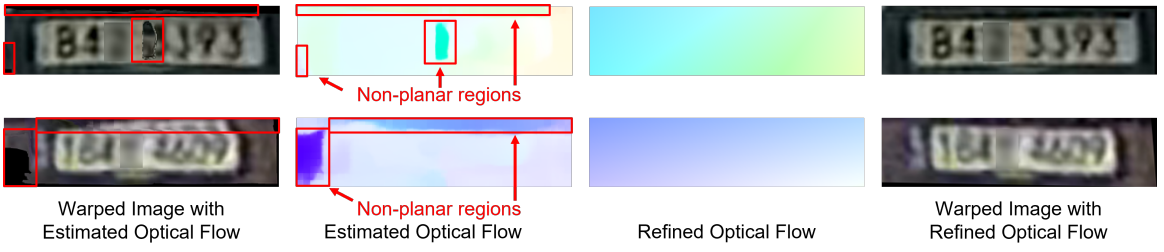

Figure 10: Qualitative results of the Spatial Refinement Module in MF-LPR². Our module effectively refines erroneous optical flows by removing non-planar regions, achieving an additional improvement of 10.47%p in LPR accuracy.

filtering and spatial refinement modules. The second variant includes the temporal filtering module but excludes the spatial refinement module. These variants allow us to dissect the impact of each module on key metrics, such as PSNR, SSIM, and License Plate Recognition (LPR) accuracy.

Table 5 presents the evaluation results of the three model configurations. Adding the temporal filtering module resulted in improvements in PSNR (from 16.3114 to 16.3240), SSIM (from 0.3468 to 0.3480) and increased LPR accuracy by 1.27%p (i.e, 1.27 percentage points, from 74.70% to 75.97%). Moreover, incorporating both the temporal filtering and spatial refinement modules significantly improved performance across all metrics, achieving the highest

LPR accuracy of 86.44%, an increase of 11.74%p compared to the baseline model without the modules. This result demonstrates the complementary nature of the two modules, where temporal filtering enhances inter-frame consistency and spatial refinement fine-tunes localized errors, leading to substantial performance gains.

To further validate the effectiveness of each module, we provide qualitative visualizations. Figure 9 illustrates how the temporal filtering module reduces misaligned flows by rejecting poorly estimated optical flows, while Figure 10 highlights the spatial refinement module's ability to improve the accuracy of optical flows by refining non-planar regions.

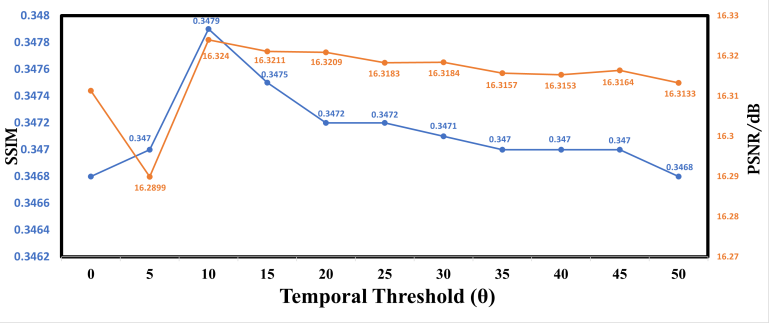

Figure 11: The performance of MF-LPR² by the threshold value θ_{temp} for the temporal filtering module.

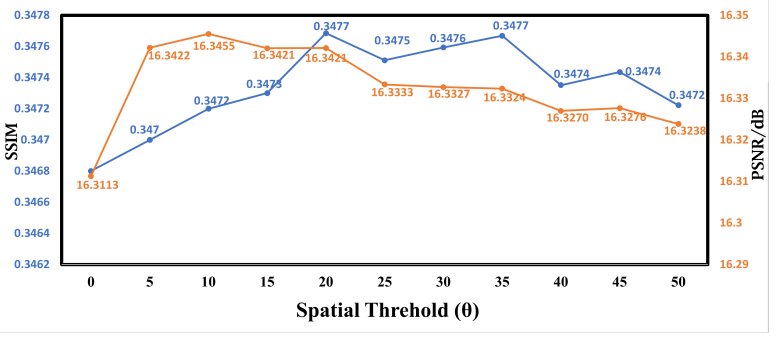

Figure 12: The performance of MF-LPR² by the threshold value $\theta_{spatial}$ for the spatial refinement module.

4.6.3. Threshold Values for Temporal Filtering and Spatial Refinement

To find the optimal values for θ_{temp} and $\theta_{spatial}$, we measured SSIM and PSNR by changing the threshold values. Fig. 11 and Fig. 12 present the results of ablation studies on θ_{temp} and $\theta_{spatial}$, respectively. In Fig. 11, the highest SSIM value of 0.3479 and PSNR value of 16.324 are observed at a temporal threshold of 10. Beyond this threshold value, both SSIM and PSNR values gradually decrease. The reason of the declination is that, as the temporal threshold increases, fewer images are filtered, leaving more optical flow errors. These errors negatively impact the structural similarity and image quality, thereby causing a reduction in the SSIM and PSNR metrics. In Fig. 12, the highest SSIM and PNSR values are observed with different $\theta_{spatial}$ values. The highest SSIM value of 0.3477 is observed at threshold values of 20 and 35, while the highest PSNR value of 16.3455 is observed at a threshold of 10. Setting $\theta_{spatial} = 20$ achieves a good balance, yielding high values in both SSIM and PSNR.

5. Conclusion

In this paper, we proposed MF-LPR², a novel framework for multi-frame license plate image restoration and recognition. MF-LPR² effectively addresses the image degradation issues frequently found in low-quality dash cam images, such as insufficient resolution, motion blur, and light glare, by aggregating complementary information from

neighboring frames. Unlike most deep-learning-based super-resolution methods that synthesize a new high-quality image, MF-LPR² does not distort the content of the input image or produce spurious artifacts, thereby preserving the evidential value of the original input. In experiments, MF-LPR² outperformed eight recently developed image/video restoration models and three license plate image restoration and recognition models, by exhibiting a significantly improved character recognition rate. These results suggest that our work introduces meaningful improvements that are useful in various application fields, such as traffic law enforcement, crime investigation, and surveillance.

One limitation of MF-LPR² is that it produces output images solely based on the input images. Referring to contextual information, e.g., utilizing the feedback from the recognizer or leveraging the special characteristics of license plate images, may further improve its performance. These topics can be explored in future work.

6. Acknowledgments

This work was supported by NC& Co., Ltd., National Program for Excellence in SW supervised by the Institute of Information&Communications Technology Planning&Evaluation (IITP, Korea) supported by Ministry of Science and ICT(MSIT, Korea) in 2023 (2023-0-00055), and Artificial intelligence industrial convergence cluster development project funded by the Ministry of Science and ICT(MSIT, Korea)&Gwangju Metropolitan City.

CRediT authorship contribution statement

Kihyun Na: Conceptualization of this study, Methodology, Software, Writing, Validation, Supervision. **Junseok Oh:** Conceptualization of this study, Methodology, Software. **Youngkwan Cho:** Software, Data Curation, Validation, Writing - Original Draft. **Bumjin Kim:** Software, Data Curation, Validation, Visualization. **Sungmin Cho:** Data Curation. **Jinyoung Choi:** Validation, Writing - Review & Editing. **Injung Kim:** Conceptualization of this study, Methodology, Writing - Review & Editing, Project administration.

References

[1] C. Dong, C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 2, pp. 295-307, 1 Feb. 2016.

[2] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4681-4690.

[3] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks," *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp 0-0.

[4] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, 2017, pp. 136-144.

[5] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," Proc. Eur. Conf. Comput. Vis. (ECCV), 2018, pp. 286-301.

[6] X. Chen, X. Wang, W. Zhang, X. Kong, Y. Qiao, J. Zhou, and C. Dong, "HAT: Hybrid Attention Transformer for Image Restoration," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp.22367-22377.

[7] X. Wang, L. Xie, C. Dong, and Y. Shan, "Real-esrgan: Training real-world blind super-resolution with pure synthetic data," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 1905-1914.

[8] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a Practical Degradation Model for Deep Blind Image Super-Resolution.," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2021, pp. 4791-4800.

[9] H. Chen, Y. Wang, T. Guo, C. Xu, Y. Deng, Z. Liu, S. Ma, C. Xu, and W. Gao, "Pre-trained image processing Transformer," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 12299-12310.

[10] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, "SwinIR: Image Restoration Using Swin Transformer," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV) Workshops*, Oct. 2021, pp. 1833-1844.

[11] S. Zamir, A. Arora, S. Khan, M. Hayat, F. Khan, M. Yang, "Restormer: Efficient Transformer for high-resolution image restoration," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 5728-5739.

[12] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped Transformer for image restoration," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 17683-17693.

[13] Hsu, C., Lee, C. & Chou, Y. DRCT: Saving Image Super-resolution away from Information Bottleneck. (2024), https://arxiv.org/abs/2404.00722

[14] Luo, Z., Huang, H., Yu, L., Li, Y., Fan, H. & Liu, S. Deep Constrained Least Squares for Blind Image Super-Resolution. (2022), https://arxiv.org/abs/2202.07508

[15] K. Chan, S. Zhou, X. Xu, and C. Loy, "Basicvsr++: Improving video super-resolution with enhanced propagation and alignment," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 5972-5981.

[16] S. Shi, J. Gu, L. Xie, X. Wang, Y. Yang, and C. Dong, "Rethinking alignment in video super-resolution Transformers," *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2022, pp. 36081-36093.
[17] X. Wang, K. Chan, K. Yu, C. Dong, C. Change Loy, "Edvr: Video restoration with enhanced deformable convolutional networks," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR) Workshops*, 2019, pp. 0-0.
[18] J. Liang, J. Cao, Y. Fan, K. Zhang, R. Ranjan, Y. Li, R. Timofte, L. Van Gool, "Vrt: A video restoration Transformer," *IEEE Trans. Image Process*, vol. 33, pp. 2171-2182, 2024
[19] J. Liang, Y. Fan, X. Xiang, R. Ranjan, E. Ilg, S. Green, J. Cao, K. Zhang, R. Timofte, and L. Van Gool, "Recurrent video restoration Transformer with guided deformable attention," *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2022, pp. 378-393.
[20] Xu, K., Yu, Z., Wang, X., Mi, M. & Yao, A. "Enhancing Video Super-Resolution via Implicit Resampling-based Alignment," (2024), https://arxiv.org/abs/2305.00163
[21] W. Wang, E. Xie, P. Sun, W. Wang, L. Tian, C. Shen, and P. Luo, "Textsr: Content-aware text super-resolution guided by recognition," 2019, *arXiv:1909.07113*. [Online]. Available: https://arxiv.org/abs/1909.07113
[22] W. Wang, E. Xie, X. Liu, W. Wang, D. Liang, C. Shen, and X. Bai, "Scene text image super-resolution in the wild," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2020, pp. 650-666.
[23] J. Ma, S. Guo, and L. Zhang, "Text prior guided scene text image super-resolution," *IEEE Trans. Image Process*, 2023, pp. 1341-1353.
[24] J. Chen, B. Li, and X. Xue, "Scene text telescope: Text-focused scene image super-resolution," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2021, pp. 12026-12035.
[25] J. Ma, Z. Liang, and L. Zhang, "A text attention network for spatial deformation robust scene text image super-resolution," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2022, pp. 5911-5920.
[26] S. Zhu, Z. Zhao, P. Fang, and H. Xue, "Improving Scene Text Image Super-Resolution via Dual Prior Modulation Network," *Proc. AAAI Conf. Artif. Intell. (AAAI)*, 2023, pp. 3843-3851.
[27] D. Lucas, and T. Kanade. "An iterative image registration technique with an application to stereo vision." *7th international joint conference on Artificial intelligence(IJCAI)*, vol.2, 1981.
[28] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Hazirbas, V. Golkov, P. V. D. Smagt, D. Cremers, and T. Brox, "FlowNet: Learning Optical Flow With Convolutional Networks," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 2758-2766.
[29] A. Ranjan, and M. J. Black, "Optical flow estimation using a spatial pyramid network," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4161-4170.
[30] Z. Huang, X. Shi, C. Zhang, Q. Wang, K. C. Cheung, H. Qin, J. Dai, and H. Li, "Flowformer: A Transformer Architecture for Optical Flow," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp.668-685.
[31] X. Shi, Z. Huang, D. Li, M. Zhang, K. Cheung, S. See, H. Qin, J. Dai, and H. Li, "Flowformer++: Masked cost volume autoencoding for pretraining optical flow estimation," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2023, pp. 1599-1610.
[32] S. Ren, K. He, R. Girshick, and J Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.39, no.6, pp. 1137-1149, 2016.
[33] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2016, pp. 21-37.
[34] J. Redmon, and A. Farhadi, "YOLO9000: better, faster, stronger," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 7263-7271.
[35] W. Zhou, Q. Guo, J. Lei, L. Yu and J. -N. Hwang, "ECFFNet: Effective and Consistent Feature Fusion Network for RGB-T Salient Object Detection," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 32, no. 3, pp. 1224-1235, March 2022, doi: 10.1109/TCSVT.2021.3077058.
[36] W. Zhou, Y. Zhu, J. Lei, J. Wan and L. Yu, "CCAFNet: Crossflow and Cross-Scale Adaptive Fusion Network for Detecting Salient Objects in RGB-D Images," in IEEE Transactions on Multimedia, vol. 24, pp. 2192-2204, 2022, doi: 10.1109/TMM.2021.3077767.
[37] W. Zhou, Y. Zhu, J. Lei, J. Wan and L. Yu, "APNet: Adversarial Learning Assistance and Perceived Importance Fusion Network for All-Day RGB-T Salient Object Detection," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 4, pp. 957-968, Aug. 2022, doi: 10.1109/TETCI.2021.3118043.
[38] W. Zhou, F. Sun, Q. Jiang, R. Cong and J. -N. Hwang, "WaveNet: Wavelet Network With Knowledge Distillation for RGB-T Salient Object Detection," in IEEE Transactions on Image Processing, vol. 32, pp. 3027-3039, 2023, doi: 10.1109/TIP.2023.3275538.
[39] W. Zhou, Y. Zhu, J. Lei, R. Yang and L. Yu, "LSNet: Lightweight Spatial Boosting Network for Detecting Salient Objects in RGB-Thermal Images," in IEEE Transactions on Image Processing, vol. 32, pp. 1329-1340, 2023, doi: 10.1109/TIP.2023.3242775.
[40] Silva, S. & Jung, C. License plate detection and recognition in unconstrained scenarios. *Proceedings Of The European Conference On Computer Vision (ECCV)*. pp. 580-596 (2018)
[41] J. Sun, W. Cao, Z. Xu, and J. Ponce, "Learning a Convolutional Neural Network for Non-uniform Motion Blur Removal," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 769-777.
[42] P. Svoboda, M. Hradiš, L. Maršík, and P. Zemcík, "CNN for license plate motion deblurring," *IEEE Int. Conf. Image Process. (ICIP)*, 2016, pp. 3832-3836.
[43] Q. Lu, W. Zhou, L. Fang, and H. Li, "Robust blur kernel estimation for license plate images from fast moving vehicles," *IEEE Trans. Image Process.*, vol. 25, no. 5, pp. 2311-2323, 2016.
[44] P. Rao, and R. Muthu, "A new de-blurring technique for license plate images with robust length estimation," *Int. Conf. Intell. Comput. Control (I2C2)*, 2017, pp. 1-6.
[45] V. Nguyen, and D. Nguyen, "Joint image deblurring and binarization for license plate images using deep generative adversarial networks," *NAFOSTED Conf. Inf. Comput. Sci. (NICS)*, 2018, pp. 430-435.
[46] O. Kupyn, V. Budzan, M. Mykhailych, D. Mishkin, and J. Matas, "DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 8183-8192.

[47] Y. Zou, Y. Wang, W. Guan, and W. Wang, “Semantic super-resolution for extremely low-resolution vehicle license plate," *IEEE Int. Conf. Acoust, Speech, Signal Process. (ICASSP)*, 2019, pp. 3772-3776.

[48] V. Nascimento, R. Laroca, J. Lambert, W. Schwartz, and D. Menotti, “Super-resolution of license plate images using attention modules and sub-pixel convolution layers," *Comput. & Graph.*, vol. 113, pp. 69-76, 2023.

[49] Kim, D., Kim, J. & Park, E. AFA-Net: Adaptive Feature Attention Network in image deblurring and super-resolution for improving license plate recognition. *Computer Vision And Image Understanding*. pp. 238. 103879 (2024).

[50] H. Seibel, S. Goldenstein, and A. Rocha, “Eyes on the target: Super-resolution and license-plate recognition in low-quality surveillance videos.," *IEEE Access*, vol. 5, pp. 20020-20035, 2017

[51] Zhou Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, “Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, 2004

[52] R. Zhang, P. Isola, A. A. Efros, E. Shechtman and O. Wang, “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, 2018, pp. 586-595.

[53] L. C. Chen, G. Papandreou, F. Schroff, H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*. [Online]. Available:https://arxiv.org/abs/1706.05587

[54] Dbpprt, “pytorch-licenseplate-segmentation," *github.com*. Accessed: May. 6, 2024. [Online]. Available: https://github.com/dbpprt/pytorch-licenseplate-segmentation

[55] XiaoyuShi97, “FlowFormerPlusPlus," *github.com*. Accessed: May. 6, 2024. [Online]. Available: https://github.com/XiaoyuShi97/FlowFormerPlusPlus

[56] H. Seibel, S. Goldenstein, and A. Rocha, “Fast and Effective Geometric K-Nearest Neighbors Multi-frame Super-Resolution," *Proc. Conf. Graph. Patterns Images (SIBGRAPI)*, 2015, pp. 103-110.

[57] P. Wang, C. Da, and C. Yao, “Multi-granularity prediction for scene text recognition," *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2022, pp. 339-355.

[58] AlibabaResearch, “AdvancedLiterateMachinery," *github.com* Accessed: May. 6, 2024. [Online]. Available: https://github.com/AlibabaResearch/AdvancedLiterateMachinery

[59] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, “Toward Real-World Single Image Super-Resolution: A New Benchmark and A New Model," *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2019, pp. 13233-13242.

[60] Openalpr, “OpenALPR-EU," *github.com* Accessed: May. 6, 2024. [Online]. Available: https://github.com/openalpr/benchmarks/tree/master/endtoend/eu

[61] G. Gonçalves, S. Silva, D. Menotti, and W. Schwartz, “Benchmark for license plate character segmentation," *Electron. Imag.*, vol.25, Oct. 2016, dol:http://dx.doi.org/10.1117/1.JEI.25.5.053034.

[62] R. Laroca, E. Severo, L. Zanlorensi, L. Oliveira, G. Gonçalves, W. Schwartz, and D. Menotti, “A Robust Real-Time Automatic License Plate Recognition Based on the YOLO Detector," *Int. Joint Conf. Neural Networks (IJCNN)*, 2018, pp. 1-10.

[63] V. Nascimento, R. Laroca, J. Lambert, W. Schwartz, and D. Menotti, “Combining Attention Module and Pixel Shuffle for License Plate Super-Resolution," *Proc. Conf. Graph. Patterns Images (SIBGRAPI)*, 2022, pp. 228-233.