

Article

DiffPlate: A Diffusion Model for Super-Resolution of License Plate Images

Sawsan AlHalawani , Bilel Benjdira * , Adel Ammar , Anis Koubaa  and Anas M. Ali 

Robotics and Internet-of-Things Lab, Prince Sultan University, Riyadh 11586, Saudi Arabia; shalawani@psu.edu.sa (S.A.); aammar@psu.edu.sa (A.A.); akoubaa@psu.edu.sa (A.K.); aaboessa@psu.edu.sa (A.M.A.)

* Correspondence: bbenjdira@psu.edu.sa

Abstract: License plate recognition is a pivotal challenge in surveillance applications, predominantly due to the low resolution and diminutive size of license plates, which impairs recognition accuracy. The advent of AI-based super-resolution techniques offers a promising avenue to ameliorate the resolution of such images. Despite the deployment of various super-resolution methodologies, including Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs), the quest for satisfactory outcomes in license plate image enhancement persists. This paper introduces “DiffPlate”, a novel Diffusion Model specifically tailored for license plate super-resolution. Leveraging the unprecedented capabilities of Diffusion Models in image generation, DiffPlate is meticulously trained on a dataset comprising low-resolution and high-resolution pairs of Saudi license plates, curated for our surveillance application. Our empirical analysis substantiates that DiffPlate markedly eclipses state-of-the-art alternatives such as SwinIR and ESRGAN, evidencing a 26.47% and 37.32% enhancement in Peak Signal-to-Noise Ratio (PSNR) against these benchmarks, respectively. Furthermore, DiffPlate achieves superior performance in terms of Structural Similarity Index (SSIM), with a 4.88% and 16.21% improvement over SwinIR and ESRGAN, respectively. Human evaluative studies further corroborate that images refined by DiffPlate were preferred 92% more frequently compared to those processed by other algorithms. Through DiffPlate, we present a new solution to the license plate super-resolution challenge, demonstrating significant potential for adoption in real-world surveillance systems.

Keywords: DiffPlate; image super-resolution; diffusion model; license plate super-resolution; computer vision



Citation: AlHalawani, S.; Benjdira, B.; Ahmed, M.; Koubaa, A.; Ali, A.M. DiffPlate: A Diffusion Model for Super-Resolution of License Plate Images. *Electronics* **2024**, *13*, 2670. <https://doi.org/10.3390/electronics13132670>

Academic Editor: George A. Tsihrintzis

Received: 30 May 2024

Revised: 2 July 2024

Accepted: 2 July 2024

Published: 7 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

As the number of motor vehicles continues to rise, license plates become crucial to public security [1]. Computer vision techniques offer timely solutions for license plate identification. A prevalent technique is License Plate Recognition (LPR) [2–7], which transforms license plate images into recognizable character sets. This aids in tasks such as vehicle tracking for security, traffic violations monitoring [8–11], toll collection [12], and intelligent parking [13,14].

The effectiveness of LPR hinges on the quality of its input images. While surveillance cameras capture these images, many output images of subpar quality and low resolution. Factors such as camera positioning, weather, and lighting conditions play roles in this degradation. Given that license plates occupy a minor section of these images, they often appear at even lower resolutions. Enhancing and denoising these images is vital for practical applications and remains an active research area [15–20].

Super-resolution (SR) is a reconstruction technology that produces high-resolution images from low-resolution images. Super-resolution has gained researchers’ attention recently. The first paper that addressed this problem was by Dong et al. [21]. They defined

the mapping between low- and high-resolution images using a Convolutional Neural Network (CNN). They demonstrated good restoration quality that was fast to achieve. Since then, many research studies have been published in the area with better results.

There has been some research work that addressed enhancing LP images using SR techniques. Some authors used older methods such as image processing and interpolation techniques [22]. Such approaches depend on the pixel information, which may lead to poor quality. Another approach utilized multiple low-resolution images to produce a single high-resolution image [23]. Other approaches [24–26] utilized the SRGAN model to produce high-quality images. Figure 1 shows some examples where traditional methods could fail to produce images which mimic the original image exactly with high quality.

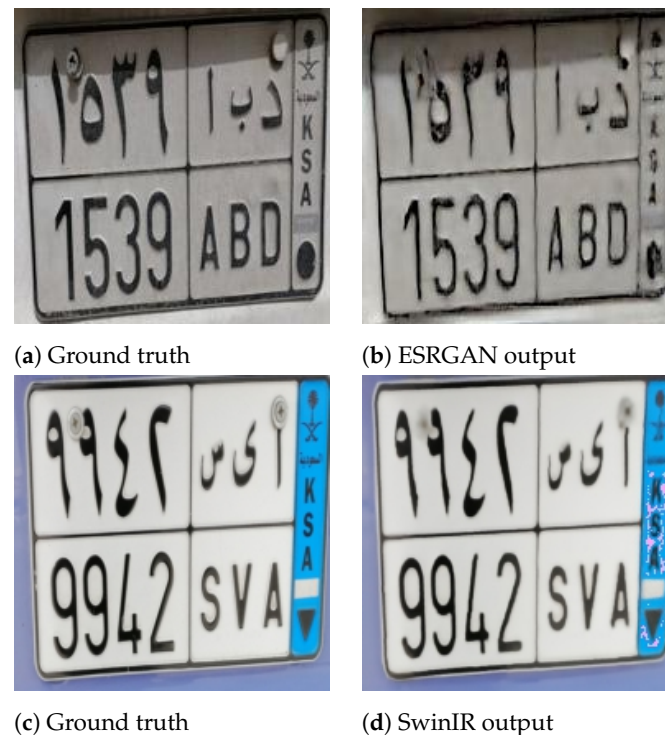


Figure 1. Comparison between the ground truth and the super-resolved images using SwinIR and ESRGAN.

Recently, the Denoising Diffusion Probabilistic Model (DDPM) [27] has garnered significant interest, often surpassing alternatives like GAN-based methods. The DDPM's superior performance is evident across diverse domains, such as image generation [28], super-resolution [29,30], deblurring [31], and more. Notably, while the DDPM predominantly addresses general images, its application for license plate super-resolution remains unexplored.

In this study, we introduce DiffPlate, a pioneering adaptation of Diffusion Models for License Plate (LP) Image Super-Resolution. Our dataset comprises genuine Saudi Arabia license plate images. Through comprehensive experiments, we benchmarked the DiffPlate approach against leading super-resolution methods like SwinIR [32] and ESRGAN [33], employing PSNR, SSIM, and MS-SSIM as evaluation metrics. Notably, DiffPlate enhanced PSNR values by 13% and 37% over SwinIR and ESRGAN, respectively, with SSIM improving by 5% and 18%. A human-centric evaluation confirmed our method's efficacy: 92% of participants favored our generated images over those of SwinIR and ESRGAN. This underscores the potential of Diffusion Models in reconstructing high-fidelity LP images, indicating an exciting avenue for future research.

The primary contributions of this study are as follows:

- Introducing DiffPlate, a novel method for License Plate Image Super-Resolution, demonstrating superior performance over existing state-of-the-art approaches.

- Pioneering the application of Diffusion Models for License Plate Image Super-Resolution, to the best of our knowledge.
- Development of a unique dataset focused on Saudi Arabian license plates.

Section 2 presents a summary of related work in license plate image SR. Our methodology is elucidated in Section 3. Comprehensive experimental results, both quantitative and qualitative, are discussed in Section 4. This paper concludes with discussions in Section 5 and final remarks in Section 6.

2. Related Works

This section reviews prior research pertinent to License Plate Image Super-Resolution.

Image super-resolution (ISR) aims to produce high-resolution images from their lower-resolution counterparts. Although ISR has wide applications, this paper focuses on license plates. Recent methodologies predominantly employ deep learning, especially Convolutional Neural Networks (CNNs) and Generative Adversarial Networks (GANs).

Yang et al. [34] introduced a multi-scale super-resolution CNN (MSRCNN), drawing inspiration from GoogLeNet [35]. Their model has three convolutional layers designed for feature extraction, feature mapping, and high-resolution image reconstruction. Importantly, they found that training with license plate datasets results in superior super-resolution specifically for license plate images. Lee et al. [25] built upon the SRGAN architecture and incorporated a character-based perceptual loss for high-quality image generation suitable for Optical Character Recognition (OCR). Their method, when compared with the traditional SRGAN, improved OCR accuracy by 6.8% and 4.0% at the plate and character levels, respectively.

Hamdi et al. [26] developed a Double GAN for image enhancement and super-resolution. While their model exhibited significant image quality improvement, it was computationally intensive. Lin et al. [30] targeted Chinese license plate images enhancement using a GAN. They introduced a residual dense network and progressive up-sampling, achieving better performance metrics and reduced reconstruction time compared to several benchmarks. Shah et al. [36] employed the VGG-19 network in conjunction with a GAN for super-resolution, effectively addressing the over-smoothing issue. Their approach demonstrated a marked improvement in PSNR and accuracy. Nascimento et al. [37] extended the Multi-Path Residual Network (MPRNet) [38], merging PixelShuffle layers with attention modules. Their model, designed to enhance license plate images, outperformed the baseline MPRNet in various aspects. Lee et al. [39] proposed a two-step framework, first training an ISR network, followed by a character recognition network. While their global image extraction technique showed promise, it faltered under challenging conditions like low lighting and rotations.

In another contribution, ref. [40] combined ISR with feature extraction for license plate recognition. Their approach optimized weight freezing techniques to leverage high-frequency image details and character information, albeit with extended training durations. Our research pioneers the adaptation of Diffusion Models [27] for LP super-resolution. To date, Diffusion Models, while utilized for generic image denoising, have not been specifically applied to LP images. Notably, Ho et al. [41] and Saharia et al. [29] showcased their potential in denoising and super-resolution for general and remote sensing images, respectively. A summary of the discussed research can be found in Table 1.

Table 1. A summary of the related work which is addressed in this paper.

Ref.	Key Contributions	LP Dataset	PSNR dB	SSIM	Year
[34]	Multi-scale super-resolution CNN (MSRCNN).	Datasets are collected by video surveillance systems.	32.14	0.936	2018
[25]	Character-based perceptual loss with SRGAN	Created their dataset from public LP datasets [42].	-	-	2020
[26]	Double GAN network (D_GAN_ESR)	Custom two datasets: The first added a motion blur noise. The second used a CycleGAN.	29.558	0.227	2021
[30]	Used residual connections with progressive up-sampling	Chinese City Parking Dataset (CCPD) [43]	26.08	0.77	2021
[36]	Pre-trained VGG-19	LP Images	28.69	-	2022
[37]	Single Image Reconstruction model with MPRNet and attention modules	Their created LP image dataset was released for the public at [37]	26.4	0.89	2022
[39]	SR network with character recognition network using global image extraction technique	11,428 images in the training set and 1999 images in the validation set.	34.13	-	2022
[40]	Weight freezing technique, high-frequency feature extraction	UFPR and Greek vehicle datasets	20.6	-	2022
[41]	Diffusion probabilistic models and denoising score matching with Langevin dynamics	CIFAR10 dataset	-	-	2020
[29]	Adapts denoising diffusion probabilistic models via repeated refinements.	Training face SR on Flickr-Faces-HQ (FFHQ).	23.04	0.65	2022
[44]	Proposed the generative Diffusion Model with Detail Complement (DMDC) into Remote Sensing Super-Resolution task	Potsdam (Germany) dataset and Vaihingen (Germany) dataset for remote sensing	23.46	0.6696	2022

3. Proposed Method: DiffPlate

In this section, we introduce DiffPlate, which is dedicated to the super-resolution of license plate images.

Image super-resolution aims to generate high-resolution (HR) images from their low-resolution (LR) counterparts. Such a dataset consists of paired LR and HR images, represented as follows:

$$D = \{x_i^{lr}, z_i^{hr}\}_{i=1}^N \quad (1)$$

Here, x^{lr} is the degraded LR image with z^{hr} as its original HR version. The degradation typically stems from introducing noise to the HR image or reducing its resolution, as described by the following:

$$x^{lr} = G(z^{hr}; \theta) \quad (2)$$

where θ denotes degradation parameters. The super-resolution task seeks to reconstruct y^{hr} , an approximation of the HR image z^{hr} from x^{lr} , using the following model:

$$y^{hr} = F(x^{lr}; \Phi) \quad (3)$$

While many current techniques rely on pixel-level information, this often sacrifices important image characteristics. Recognizing the effectiveness of the Diffusion Model in image generation, given its proficiency in capturing global features, we employ it for License Plate Image Super-Resolution.

Diffusion Models, rooted in Non-Equilibrium Thermodynamics, comprise two phases: the forward noise introduction and the backward noise removal processes. The former

uses a Markov chain approach to gradually degrade the image by adding Gaussian noise, visualized in Figure 2.



Figure 2. The forward process for adding noise until a completely noisy image is obtained.

As displayed in Figure 2, the forward process starts from an HR license plate image z_0^{hr} , which is either noise-introduced or downsampled to produce x_0^{lr} . Then, the forward process further adds noise until isotropy is reached at step T .

3.1. Forward Noise Introduction Process.

From a given input x_0^{lr} , each Markov chain step introduces spherical Gaussian noise with variance β_t . At step t , x_t^{lr} is formed by adding noise to x_{t-1}^{lr} as follows:

$$q(x_t^{lr}|x_{t-1}^{lr}) = \mathcal{N}(x_t^{lr}; \mu_t = \sqrt{1 - \beta_t}x_{t-1}^{lr}, \Sigma_t = \beta_t I) \quad (4)$$

This chain produces a sequence from x_0^{lr} to x_T^{lr} defined by

$$q(x_{1:T}^{lr}|x_0^{lr}) = \prod_{t=1}^T q(x_t^{lr}|x_{t-1}^{lr}) \quad (5)$$

Using reparametrization, we can succinctly describe x_t^{lr} without multiple iterations:

$$\begin{aligned} x_t^{lr} &= \sqrt{1 - \beta_t}x_{t-1}^{lr} + \sqrt{\beta_t}\epsilon_{t-1} \\ &= \sqrt{\alpha_t}x_{t-2}^{lr} + \sqrt{1 - \alpha_t}\epsilon_{t-1} \\ &= \dots \\ &= \sqrt{\bar{\alpha}_t}x_0^{lr} + \sqrt{1 - \bar{\alpha}_t}\epsilon_0 \end{aligned} \quad (6)$$

Here, α_t and $\bar{\alpha}_t$ are defined, giving us the ability to determine x_t^{lr} at any step based on β_t . Consequently, x_t^{lr} is produced by the following:

$$x_t^{lr} \sim q(x_t^{lr}|x_0^{lr}) = \mathcal{N}(x_t^{lr}; \sqrt{\bar{\alpha}_t}x_0^{lr}, (1 - \bar{\alpha}_t)I) \quad (7)$$

The backward process, as displayed in Figure 3, reconstructs the HR image y_0^{hr} from the fully degraded image x_T^{lr} , using prior knowledge about x_0^{lr} .

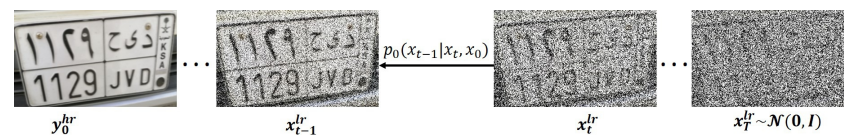


Figure 3. The backward process for reconstructing the HR image.

3.2. Backward Denoising Process

The output of the previous step produces a set of noisy images x_t^{lr} with an isotropic distribution. Our target at this stage is to recover the high-resolution LP image from the noisy image. This is achieved by reversing the previous process and gradually removing the noise from the x_t^{lr} sample, which is drawn from the normal distribution $q(x_0^{lr})$ to obtain

a sample from the original data distribution. The backward process is shown in Figure 3. This can be approximated with the parameterized model p_0 that uses a CNN as follows:

$$p_0(x_{t-1}^{lr}|x_t^{lr}) = \mathcal{N}(x_{t-1}^{lr}; \mu_\theta(x_t^{lr}, t), \Sigma_\theta(x_t^{lr}, t)) \quad (8)$$

This model $p(x_0^{lr})$ can be used to predict the $\mu_\theta(x_t^{lr}, t)$ and $\Sigma_\theta(x_t^{lr}, t)$ parameters of the Gaussian distribution at each time step t . Then, the trajectory from x_T^{lr} to x_0^{lr} can be found as follows:

$$p_0(x_{0:T}^{lr}) = p_0(x_T^{lr}) \prod_{t=1}^T p_0(x_{t-1}^{lr}|x_t^{lr}) \quad (9)$$

In order for the generative model to recover the high-resolution LP image, it needs to have some information about the original image x_0^{lr} . Therefore, we can condition the sampling of x_t^{lr} at a time step t based on x_0^{lr} . This can be represented as follows:

$$q(x_{t-1}^{lr}|x_t^{lr}, x_0^{lr}) = \mathcal{N}(x_{t-1}^{lr}; \tilde{\mu}(x_t^{lr}, x_0^{lr}), \tilde{\beta}_t I) \quad (10)$$

$$\tilde{\beta}_t = \frac{1 - \bar{\alpha}_{t-1}}{1 - \bar{\alpha}_t} \cdot \beta_t \quad (11)$$

$$\tilde{\mu}_t(x_t^{lr}, x_0^{lr}) = \frac{\sqrt{\bar{\alpha}_{t-1}}\beta_t}{1 - \bar{\alpha}_t} x_0^{lr} + \frac{\sqrt{\bar{\alpha}_t}(1 - \bar{\alpha}_{t-1})}{1 - \bar{\alpha}_t} x_t^{lr} \quad (12)$$

Therefore, we can express x_0^{lr} , given that $\epsilon \sim \mathcal{N}(0, I)$, as follows:

$$x_0^{lr} = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t^{lr} - \sqrt{1 - \bar{\alpha}_t}\epsilon) \quad (13)$$

Additionally, we can define a neural network $\epsilon_\theta(x_t^{lr}, t)$ to find the values of ϵ and the mean as follows:

$$\tilde{\mu}_\theta(x_t^{lr}, t) = \frac{1}{\sqrt{\bar{\alpha}_t}}(x_t^{lr} - \frac{\beta_t}{\sqrt{1 - \bar{\alpha}_t}}\epsilon_\theta(x_t^{lr}, t)) \quad (14)$$

Diffusion Models define the loss function as the Mean Square Error (MSE) between the added noise to the original image and the predicted noise in the backward process. We follow the same approach as Ho et al. [41] by using a simplified version of the loss function that ignores the weighting coefficients and outperforms its full original version. It can be expressed as follows:

$$L_t = \mathbb{E}_{x_0^{lr}, t, \epsilon} [||\epsilon - \epsilon_0(\sqrt{\bar{\alpha}_t}x_0^{lr} + \sqrt{1 - \bar{\alpha}_t}\epsilon, t)||^2] \quad (15)$$

DiffPlate utilizes a U-Net architecture [45] which consists of contracting and expanding paths. The contracting path is mainly responsible for increasing the image resolution. It has successive layers that capture the context using up-sampling operators. The expanding path enables precise localization by combining the high-resolution features from the contracting path with the up-sampled output. This is followed by a convolution layer that learns to generate a more precise image. The contracting path uses a large number of feature channels, which are used to pass information to the higher-resolution layers. Therefore, the expanding path will be symmetric to the contracting path, which produces a U-shaped network architecture rather than a fully connected network. Although DiffPlate utilizes the model of U-Net structure, it is customized to includes a conditioning branch at every denoising step. The input to the model includes the low-resolution license plate image, which helps to guide the model to align better with the content of the low-resolution image. Figure 4 visualizes the structure of our model.

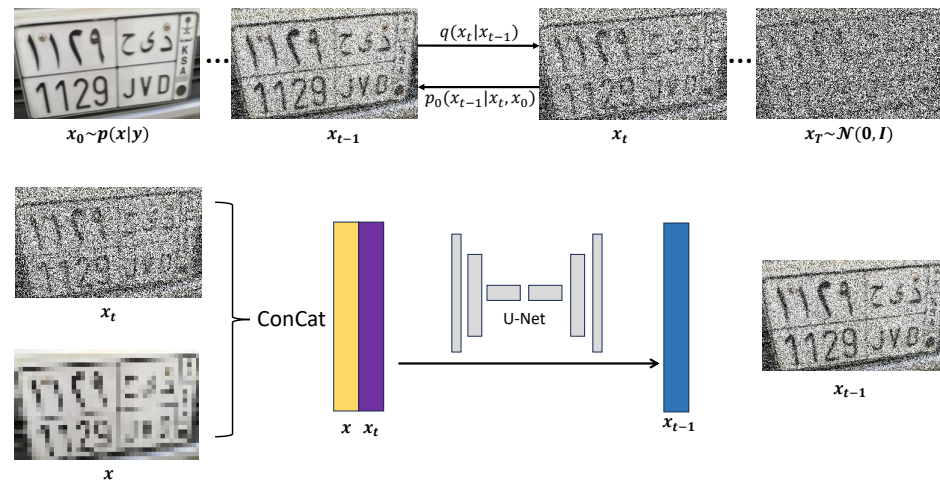


Figure 4. The structure of the proposed DiffPlate model.

4. Experiments

4.1. Datasets

Our dataset consists of genuine images of Saudi Arabian license plates (LPs). These images, captured with cameras, were meticulously cropped to focus solely on the license plates, excluding extraneous background details. Saudi license plates, featuring a mix of Arabic and English characters and numerals, represent various designs, all of which are encompassed in our study. It is essential to note that while these mixed characters might challenge certain techniques, DiffPlate sidesteps character recognition, ensuring such variations do not influence our performance.

The dataset comprises 593 color images with dimensions $192 \times 192 \times 3$, each with red (R), green (G), and blue (B) channels. For model development, we allocated 92% of the images (543 images) for training and reserved the remaining 8% (50 images) for testing. Original images underwent down-sampling by a factor of 4, resulting in low-resolution images of $48 \times 48 \times 3$. These down-sampled images were used as inputs to the Diffusion Model. For validation, DiffPlate was tested on downsampled images with identical resolution ($48 \times 48 \times 3$), emphasizing the versatility and robustness of our approach without further fine-tuning.

4.2. Evaluation Metrics

In order to assess the quality of DiffPlate comprehensively, we used the following metrics: Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM) [46], and Multi-scale Structural Similarity (MS-SSIM) [47]. PSNR represents the ratio between the maximum possible value (power) of a signal and the power of the distorting noise that degrades the quality of the image. PSNR is well suited in this context due to its simple calculations and clear meaning. It is based on using the Mean Squared Error (MSE), which compares the original pixel values to the degraded image. It can be defined as follows:

$$MSE = \frac{\sum_1^{M*N} [x - y]^2}{M * N} \quad (16)$$

$$PSNR = 10 * \log_{10} \frac{R^2}{MSE} \quad (17)$$

where R is defined based on the image format and M and N are the number of rows and columns of the image.

SSIM relies on extracting structural information from the image by assuming pixels' interdependency, which is close to the human perception. SSIM is defined based on three

characteristics which are captured from the image, i.e., luminance, contrast, and structure. They are defined as follows:

$$luminance(x, y) = \frac{2\mu_x\mu_y + \alpha_1}{\mu_x^2 + \mu_y^2 + \alpha_1} \quad (18)$$

$$contrast(x, y) = \frac{2\mu_{xy} + \alpha_2}{\sigma_x^2 + \sigma_y^2 + \alpha_2} \quad (19)$$

$$structure(x, y) = \frac{\sigma_{xy} + \alpha_3}{\sigma_x\sigma_y + \alpha_2} \quad (20)$$

where μ_x and μ_y represent the mean of x, y , respectively. σ_x and σ_y represent the standard deviation of x, y , respectively. σ_{xy} is the covariance of x, y . For division stabilization, α_1, α_2 and α_3 constants are used. A simplified representation of SSIM can be written as follows:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + \alpha_1)(\sigma_{xy} + \alpha_2)}{(\mu_x^2 + \mu_y^2 + \alpha_1)(\sigma_x^2 + \sigma_y^2 + \alpha_2)} \quad (21)$$

MS-SSIM performs multi-step down-sampling steps, providing more flexibility than SSIM and incorporating the variations in image resolution and viewing conditions. MS-SSIM can be written as follows:

$$MS-SSIM(x, y) = l_M(x, y)^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j} \quad (22)$$

where l_M is the luminance that is calculated by Equation (18) at scale M only. Moreover, c_j, s_j are the contrast and structure, which are calculated based on Equations (19) and (20), respectively. The constants $\alpha_M, \beta_j, \gamma_j$ are used to adjust the relative importance of each term.

4.3. Implementation Details

This section delineates the specifics of DiffPlate implementation and provides a comparative overview with the SwinIR and ESRGAN models.

DiffPlate leveraged Pytorch for its implementation. Conversely, SwinIR utilized Pytorch Lightning, while ESRGAN was built upon TensorFlow. For the computational demands of DiffPlate and SwinIR, we employed the NVIDIA Quadro RTX 8000 (48,601 MiB) GPU. ESRGAN, however, was implemented using the NVIDIA Tesla T4 (15,360 MiB) provided by Google Colab.

The construction of DiffPlate involved an initial step of downsizing our authentic HR images from a resolution of $192 \times 192 \times 3$ to $48 \times 48 \times 3$. To enhance the model's robustness, we augmented the dataset through random rotations at angles of [5, 10, 15]. The training regimen consisted of 8688 iterations spanning 64 epochs, with each batch comprising 4 images. Drawing from Saharia et al. [29], we used a fixed learning rate of 0.0002 and employed the Adam Optimizer with a linear warm-up schedule for our proposed model. To maintain consistency and comparability in our experiments, we applied the same fixed learning rate and optimizer settings across all models evaluated, including SwinIR and ESRGAN. This uniform approach allows for a direct comparison of model performance under equivalent training conditions.

Table 2 indicates that the Diffusion Model incurs a 21% longer training duration compared to SwinIR and 2.6% relative to ESRGAN. Nonetheless, as the ensuing section will demonstrate, the resultant performance more than compensates for this marginal increase in time.

Table 2. Training time taken by each method to train the LP super-resolution models.

	SwinIR	ESRGAN	Diffusion Model
Training Time (seconds)	9676	11,404	11,700

5. Discussion

5.1. Visualization of the Reconstruction Steps

Figure 5 shows the sample intermediate steps of the reconstruction process. It starts with the input image (a) that has 48×48 resolution. With the repetitive denoising steps following the Diffusion Model (visualized in (b)), we were able to generate a high-definition image (c) that has 192×192 resolution, which resembles the ground truth image (d). By comparing the generated HR image (c) with the ground truth image (d), we notice that it is difficult to find a difference and they are almost identical. The process illustrates the transformation from the LR image to the HR image that closely resembles the ground truth.

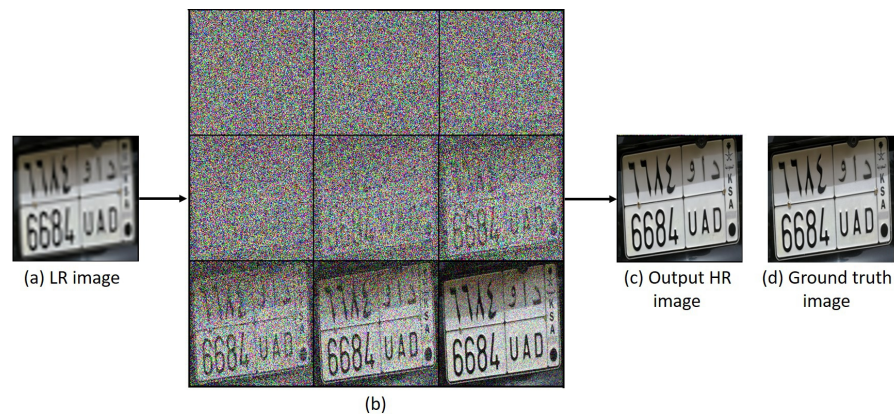


Figure 5. Intermediate steps of HR image reconstruction: (a) LR image, (b) denoised image, (c) output HR image, and (d) ground truth image.

5.2. Results

DiffPlate was evaluated on 50 images from the LP dataset. Figure 6 displays a representative high-resolution (HR) output produced using DiffPlate. These generated HR images closely match the original ground truth, preserving all image features. The resolution of the HR and generated images is (192×192) while the downsampled images are of (48×48) resolution.



Figure 6. Comparison of image resolutions: (top row) original HR image (192×192), (middle row) downsampled LR image (48×48), and (bottom row) super-resolved image (192×192) generated using the Diffusion Model.

To assess the fidelity of our generated images, we analyzed their histograms against the ground truth. Figure 7 presents histograms for the original HR image, the down-scaled low-resolution (LR) image, and the HR image generated by DiffPlate. While the LR image histogram markedly differs from the original due to noise, the histogram of the produced HR image is strikingly consistent with that of the original. This indicates DiffPlate’s capability to reproduce HR images that retain the intrinsic characteristics of the ground truth.

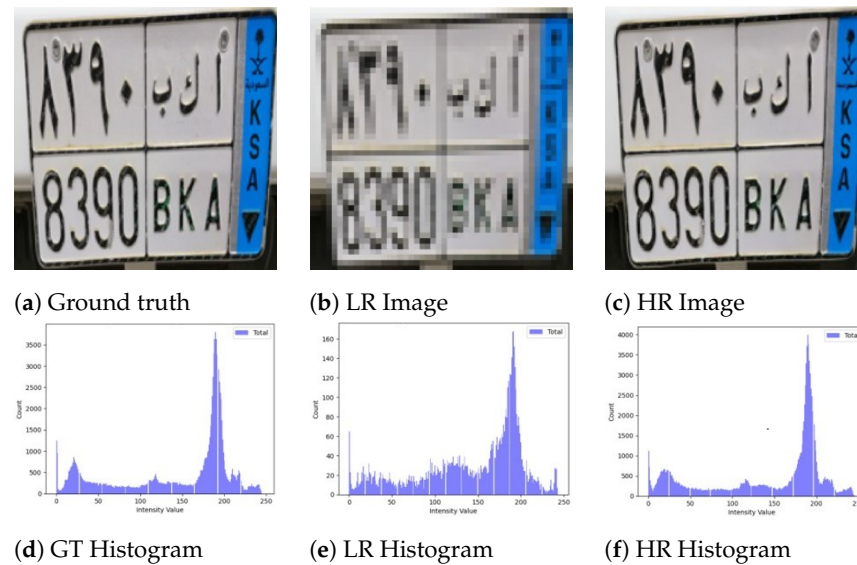


Figure 7. Sample images and their respective histograms: (top row) (a) ground truth HR image, (b) downsampled LR image, and (c) generated HR image using the Diffusion Model. (bottom row) (d) histogram of the ground truth HR image, (e) histogram of the LR image, and (f) histogram of the generated HR image, which closely resembles the ground truth histogram.

5.3. Comparison with State-of-the-Art Methods

This section discusses the comparison between our results and the results generated by two selected state-of-the-art techniques. In the recent literature, there are two main approaches which are commonly used in image generation. Some techniques use encoders and decoders to synthesize images. Others are designed based on GAN algorithms. For the first approach, we selected the state-of-the-art method SwinIR [32], while ESRGAN [33] was selected as the state-of-the-art GAN-based method. We performed quantitative and qualitative analyses, as discussed next.

5.3.1. Quantitative Results

Using Diffusion Models to generate high-resolution images in the domain of license plate image super resolution has proved to be more efficient than using the other traditional methods. We compared our approach with state-of-the-art methods, i.e., SwinIR [32] and ESRGAN [33]. We used our LP image dataset and built the SwinIR and ESRGAN models for the comparison. Table 3 shows the quantitative comparison of the LP SR task. We used three metrics for the results evaluation, i.e., PSNR, SSIM, and MS-SSIM, as discussed in Section 4.2. The results demonstrated that our approach outperformed the SOTA methods with a significant margin using the three selected evaluation metrics. First, using PSNR to assess the Diffusion Model showed a 12.55% improvement over SwinIR and a 37.32% improvement over ESRGAN. Similarly, using SSIM, DDPM outperformed SwinIR by 4.89% and ESRGAN by 17.66%. On the other hand, MS-SSIM did not show a similar very large improvement; however, it was better. This indicates that DDPM outperforms the other traditional methods in the LP SR setting.

Table 3. Quantitative results comparing our approach with SOTA methods SwinIR and ESRGAN using three metrics: PSNR, SSIM, and MS-SSIM.

Metrics	Diffusion Model	SwinIR Transformer	ESRGAN	Improvement over SwinIR	Improvement over ESRGAN
PSNR	30.6905	24.2678	22.3503	26.47%	37.32%
SSIM	0.9471	0.9030	0.8150	4.88%	16.21%
MS-SSIM	0.9934	0.9836	0.9404	0.99~1% pl	5.64%

5.3.2. Qualitative Results: Human Evaluation

While metrics such as PSNR and SSIM offer quantitative insights, they might not always align with human perception. To better understand the visual quality of images generated by the SwinIR, ESRGAN, and DiffPlate approaches, we conducted a human evaluation study.

For this, we utilized the three-alternative-forced-choice (3-AFC) discrimination test, a reputable method for subjective image quality assessment, particularly when gauging an approach's effectiveness [48]. In this paradigm, participants are presented with three options and tasked with choosing the one that best meets a given criterion.

In our study, participants answered 11 questions, each featuring images produced by the aforementioned super-resolution methods. These images were randomly ordered for each participant to eliminate potential bias. The primary directive was to identify the clearest, highest-quality image from each set. A total of 50 individuals participated.

Figure 8 reveals that over than 40 participants consistently chose our generated images as the clearest across all questions. SwinIR was preferred by only a few, and ESRGAN was the top choice for merely two questions by a single participant. Table 4 enumerates the percentage preferences for each technique, underscoring the consistent visual superiority of our generated LP images.

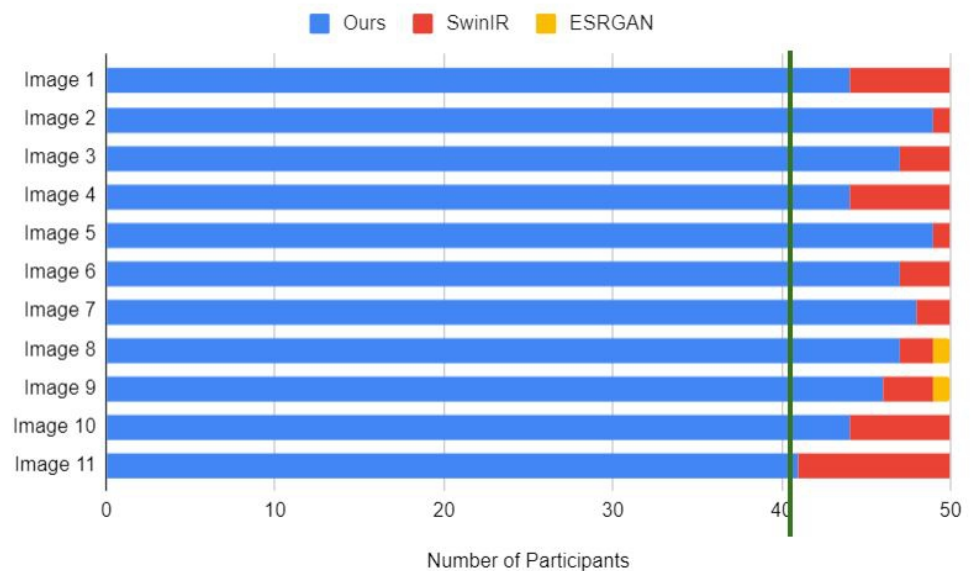
**Figure 8.** Number of participants who stated that DiffPlate generated the clearest super-resolved images.

Table 4. Participant preferences for generated LP images: 92% of participants selected the images generated by DiffPlate as the clearest. Only one participant selected the ESRGAN-generated image in two instances.

	SwinIR	ESRGAN	Diffusion Model
Average percentage of participants who selected the algorithm	8%	0%	92%

5.3.3. Qualitative Results: Visual Detail Evaluation

Upon evaluating images generated by the three algorithms, DiffPlate consistently outperformed the others. Empirical results, reinforced by human assessments, highlight the preeminence of our approach. As depicted in Figure 9, DiffPlate demonstrates superior detail restoration capabilities compared to SwinIR. The results generated by our approach outperform the SwinIR results, especially in recovering fine details, and are closer to the original images. Zoomed regions highlight that SwinIR failed to recover some details that our approach successfully restored.. Analogously, Figure 10 showcases DiffPlate’s enhanced prowess in rendering intricate details of super-resolved LP images. Zoomed regions demonstrate the efficiency of detail recovery. Our generated images have visual details that are closer to the original images compared to those generated by ESRGAN. Particularly noteworthy is the pronounced sharpness of our images, especially around character boundaries.



Figure 9. Comparison of generated images: (a) SwinIR super-resolved images, (b) images generated by DiffPlate, and (c) original images.

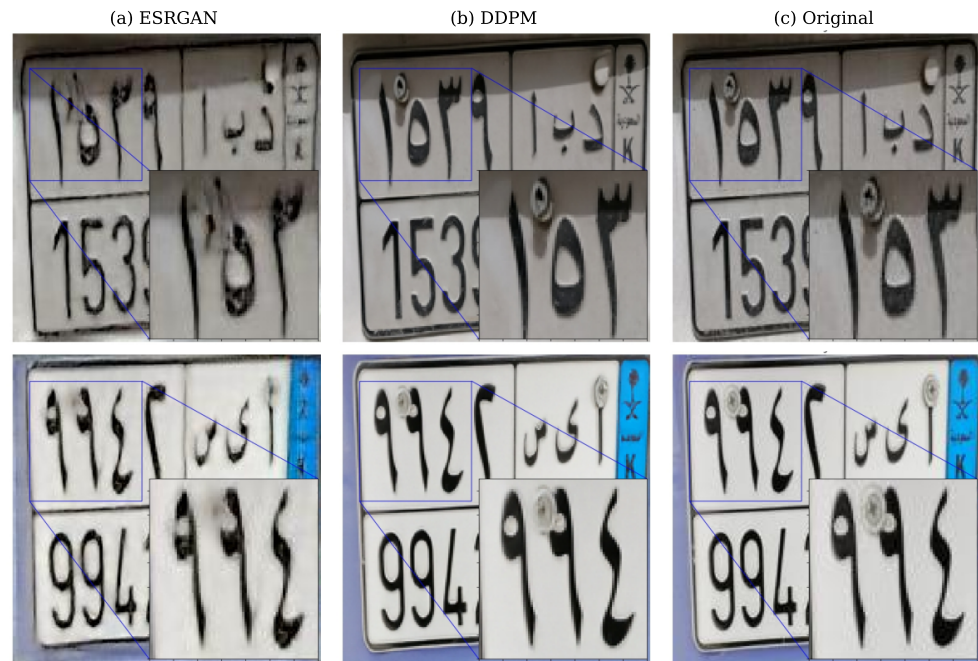


Figure 10. Comparison of generated images: (a) ESRGAN-generated images, (b) images generated by DiffPlate, and (c) original images.

6. Conclusions

In this paper, we presented a generative model that would restore a high-quality image from a highly distorted image. To the best of our knowledge, we believe that we are the first to utilize Diffusion Models in the context of License Plate Image Super-Resolution. Our results showed that our proposed approach has superior performance compared to other traditional methods. Using a Diffusion Model for LP super-resolution significantly outperforms SOTA methods such as SwinIR and ESRGAN by a notable margin. DiffPlate had a 13% PSNR improvement over SwinIR and 37% over ESRGAN. Similarly, the score of SSIM was 5% better than SwinIR and 18% better than ESRGAN. Additionally, DiffPlate was able to capture detailed features in the LP images which, on the other hand, the other SOTA methods could not achieve. According to our human evaluation experiment, 92% of our participants selected our generated images as the clearest images, against SwinIR and ESRGAN images. This shows that the Diffusion Model is a good candidate for improving the quality of license plate images, hence improving the performance of LP recognition systems. On the other hand, the main disadvantage of using the Diffusion Model approach is its heavy computational cost. This hinders its use, especially in real-time applications. However, and by considering the superior performance of the Diffusion Model, it is worth considering this approach in super-resolution problems. In the future, we will work in the direction of minimizing its computational costs while harnessing its powerful performance.

Author Contributions: Conceptualization, S.A., B.B. and A.M.A.; Methodology, S.A., B.B. and A.M.A.; Software, A.M.A.; Validation, A.M.A.; Formal analysis, S.A. and B.B.; Investigation, S.A., B.B., A.K. and A.M.A.; Resources, A.M.A.; Data curation, A.A. and A.M.A.; Writing—original draft, S.A.; Writing—review & editing, B.B., A.A., A.K. and A.M.A.; Visualization, S.A. and A.M.A.; Supervision, B.B. and A.K.; Project administration, B.B. and A.K. All authors have read and agreed to the published version of the manuscript.

Funding: The publication of this paper is funded by Prince Sultan University.

Data Availability Statement: Dataset available on request from the authors

Acknowledgments: The authors thank Prince Sultan University for their support, funding, and paying the APC for this study.

Conflicts of Interest: The authors declare no conflicts of interests associated with this publication.

References

1. Sairam, R.; Bhunia, S.S.; Thangavelu, V.; Gurusamy, M. NETRA: Enhancing IoT Security Using NFV-Based Edge Traffic Analysis. *IEEE Sens. J.* **2019**, *19*, 4660–4671. [\[CrossRef\]](#)
2. Al-Shami, S.; El-Zaart, A.; Zekri, A.; Almustafa, K.; Zantout, R. Number Recognition in the Saudi License Plates using Classification and Clustering Methods. *Appl. Math. Inf. Sci.* **2017**, *11*, 123–135. [\[CrossRef\]](#)
3. Zhuang, J.; Hou, S.; Wang, Z.; Zha, Z.J. Towards Human-Level License Plate Recognition. In Proceedings of the 15th European Conference, Munich, Germany, 14 September 2018.
4. Khan, M.A.; Sharif, M.; Javed, M.Y.; Akram, T.; Yasmin, M.; Saba, T. License number plate recognition system using entropy-based features selection approach with SVM. *IET Image Process.* **2018**, *12*, 200–209. [\[CrossRef\]](#)
5. Driss, M.; Almomani, I.; Al-Suhaimi, R.; Al-Harbi, H. Automatic Saudi Arabian License Plate Detection and Recognition Using Deep Convolutional Neural Networks. In *Proceedings of the Advances on Intelligent Informatics and Computing*; Saeed, F., Mohammed, F., Ghaleb, F., Eds.; Springer: Cham, Switzerland, 2022; pp. 3–15.
6. Ammar, A.; Koubaa, A.; Boulila, W.; Benjdira, B.; Alhabashi, Y. A multi-stage deep-learning-based vehicle and license plate recognition system with real-time edge inference. *Sensors* **2023**, *23*, 2120. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Moussaoui, H.; Akkad, N.E.; Benslimane, M.; El-Shafai, W.; Baihan, A.; Hewage, C.; Rathore, R.S. Enhancing automated vehicle identification by integrating YOLO v8 and OCR techniques for high-precision license plate detection and recognition. *Sci. Rep.* **2024**, *14*, 14389. [\[CrossRef\]](#)
8. Sikora, P.; Malina, L.; Kiac, M.; Martinasek, Z.; Riha, K.; Prinosil, J.; Jirik, L.; Srivastava, G. Artificial Intelligence-Based Surveillance System for Railway Crossing Traffic. *IEEE Sens. J.* **2021**, *21*, 15515–15526. [\[CrossRef\]](#)
9. Tahir, N.U.A.; Long, Z.; Zhang, Z.; Asim, M.; ELAffendi, M. PVswin-YOLOv8s: UAV-Based Pedestrian and Vehicle Detection for Traffic Management in Smart Cities Using Improved YOLOv8. *Drones* **2024**, *8*, 84. [\[CrossRef\]](#)
10. Alanazi, F.; Alenezi, M. Interoperability for intelligent traffic management systems in smart cities. *Int. J. Electr. Comput. Eng.* **2024**, *14*, 1864–1874.
11. Benjdira, B.; Koubaa, A.; Azar, A.T.; Khan, Z.; Ammar, A.; Boulila, W. TAU: A framework for video-based traffic analytics leveraging artificial intelligence and unmanned aerial systems. *Eng. Appl. Artif. Intell.* **2022**, *114*, 105095. [\[CrossRef\]](#)
12. Zhang, Z.; Huang, Y.; Bridgelall, R.; Al-Tarawneh, M.; Lu, P. Optimal System Design for Weigh-In-Motion Measurements Using In-Pavement Strain Sensors. *IEEE Sens. J.* **2017**, *17*, 7677–7684. [\[CrossRef\]](#)
13. Zhu, H.; Feng, S.; Yu, F. Parking Detection Method Based on Finite-State Machine and Collaborative Decision-Making. *IEEE Sens. J.* **2018**, *18*, 9829–9839. [\[CrossRef\]](#)
14. Benjdira, B.; Koubaa, A.; Boulila, W.; Ammar, A. Parking analytics framework using deep learning. In Proceedings of the 2022 2nd International Conference of Smart Systems and Emerging Technologies (SMARTTECH), Riyadh, Saudi Arabia, 9–11 May 2022; pp. 200–205.
15. Al-Shamasneh, A.R.; Ibrahim, R.W. Image Denoising Based on Quantum Calculus of Local Fractional Entropy. *Symmetry* **2023**, *15*, 396. [\[CrossRef\]](#)
16. Ali, A.M.; Benjdira, B.; Koubaa, A.; El-Shafai, W.; Khan, Z.; Boulila, W. Vision transformers in image restoration: A survey. *Sensors* **2023**, *23*, 2385. [\[CrossRef\]](#) [\[PubMed\]](#)
17. Ali, A.M.; Benjdira, B.; Koubaa, A.; Boulila, W.; El-Shafai, W. TESR: Two-Stage Approach for Enhancement and Super-Resolution of Remote Sensing Images. *Remote Sens.* **2023**, *15*, 2346. [\[CrossRef\]](#)
18. Benjdira, B.; Koubaa, A.; Ali, A.M. ROSGPT_Vision: Commanding Robots Using Only Language Models' Prompts. *arXiv* **2023**, arXiv:2308.11236.
19. Benjdira, B.; Ali, A.M.; Koubaa, A. Guided Frequency Loss for Image Restoration. *arXiv* **2023**, arXiv:2309.15563.
20. Benjdira, B.; Ali, A.M.; Koubaa, A. Streamlined Global and Local Features Combinator (SGLC) for High Resolution Image Dehazing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 17–24 June 2023; pp. 1854–1863.
21. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [\[CrossRef\]](#)
22. Ghoneim, M.; Rehan, M.; Othman, H. Using super resolution to enhance license plates recognition accuracy. In Proceedings of the 2017 12th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 19–20 December 2017; pp. 515–518.
23. Guarnieri, G.; Fontani, M.; Guzzi, F.; Carrato, S.; Jerian, M. Perspective registration and multi-frame super-resolution of license plates in surveillance videos. *Forensic Sci. Int. Digit. Investig.* **2021**, *36*, 301087. [\[CrossRef\]](#)
24. Zhang, M.; Liu, W.; Ma, H. Joint License Plate Super-Resolution and Recognition in One Multi-Task Gan Framework. In Proceedings of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, Canada, 15–20 April 2018; pp. 1443–1447.
25. Lee, S.; Kim, J.H.; Heo, J.P. Super-Resolution of License Plate Images via Character-Based Perceptual Loss. In Proceedings of the 2020 IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 560–563.

26. Hamdi, A.; Chan, Y.K.; Koo, V.C. A New Image Enhancement and Super Resolution technique for license plate recognition. *Heliyon* **2021**, *7*, e08341. [\[CrossRef\]](#)
27. Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; Ganguli, S. Deep Unsupervised Learning using Nonequilibrium Thermodynamics. In Proceedings of the 32nd International Conference on Machine Learning, Lille, France, 7–9 July 2015; Volume 37, pp. 2256–2265.
28. Dhariwal, P.; Nichol, A. Diffusion Models Beat GANs on Image Synthesis. *Adv. Neural Inf. Process. Syst.* **2021**, *34*, 8780–8794.
29. Saharia, C.; Ho, J.; Chan, W.; Salimans, T.; Fleet, D.J.; Norouzi, M. Image Super-Resolution Via Iterative Refinement. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4713–4726. [\[CrossRef\]](#)
30. Lin, M.; Liu, L.; Wang, F.; Li, J.; Pan, J. License Plate Image Reconstruction Based on Generative Adversarial Networks. *Remote Sens.* **2021**, *13*, 3018. [\[CrossRef\]](#)
31. Whang, J.; Delbracio, M.; Talebi, H.; Saharia, C.; Dimakis, A.G.; Milanfar, P. Deblurring via Stochastic Refinement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 16293–16303.
32. Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Van Gool, L.; Timofte, R. SwinIR: Image Restoration Using Swin Transformer, In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021.
33. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Loy, C.C.; Qiao, Y.; Tang, X. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
34. Yang, Y.; Bi, P.; Liu, Y. License Plate Image Super-Resolution Based on Convolutional Neural Network. In Proceedings of the 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC), Chongqing, China, 27–29 June 2018; pp. 723–727.
35. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014.
36. Shah, B.K.; Yadav, A.; Dixit, A. License Plate Image Super Resolution Using Generative Adversarial Network(GAN). In Proceedings of the 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), Salem, India, 9–11 May 2022; pp. 1139–1143.
37. Nascimento, V.; Laroca, R.; Lambert, J.d.A.; Schwartz, W.R.; Menotti, D. Combining Attention Module and Pixel Shuffle for License Plate Super-Resolution. In Proceedings of the 2022 35th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI), Natal, Brazil, 24–27 October 2022; Volume 1, pp. 228–233.
38. Mehri, A.; Ardakani, P.B.; Sappa, A.D. MPRNet: Multi-Path Residual Network for Lightweight Image Super Resolution. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020.
39. Lee, S.; Yun, J.; Yoo, S.B. Alternative Collaborative Learning for Character Recognition in Low-Resolution Images. *IEEE Access* **2022**, *10*, 22003–22017. [\[CrossRef\]](#)
40. Lee, S.J.; Yun, J.S.; Lee, E.J.; Yoo, S.B. HIFA-LPR: High-Frequency Augmented License Plate Recognition in Low-Quality Legacy Conditions via Gradual End-to-End Learning. *Mathematics* **2022**, *10*, 1569. [\[CrossRef\]](#)
41. Ho, J.; Jain, A.; Abbeel, P. Denoising Diffusion Probabilistic Models. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 6840–6851.
42. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-Time Single Image and Video Super-Resolution Using an efficient Sub-Pixel Convolutional Neural Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016.
43. Xu, Z.; Yang, W.; Meng, A.; Lu, N.; Huang, H.; Ying, C.; Huang, L. Towards End-to-End License Plate Detection and Recognition: A Large Dataset and Baseline. In *Proceedings of the Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Cham, Switzerland, 2018; pp. 261–277.
44. Liu, J.; Yuan, Z.; Pan, Z.; Fu, Y.; Liu, L.; Lu, B. Diffusion Model with Detail Complement for Super-Resolution of Remote Sensing. *Remote Sens.* **2022**, *14*, 4834. [\[CrossRef\]](#)
45. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, 5–9 October 2015*; Springer International Publishing: Cham, Switzerland, 2015.
46. Wang, Z.; Bovik, A.; Sheikh, H.; Simoncelli, E. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [\[CrossRef\]](#)
47. Wang, Z.; Simoncelli, E.; Bovik, A. Multiscale structural similarity for image quality assessment. In Proceedings of the Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 9–12 November 2003; Volume 2, pp. 1398–1402.
48. Verdun, F.; Racine, D.; Ott, J.; Tapiovaara, M.; Toroi, P.; Bochud, F.; Veldkamp, W.; Schegerer, A.; Bouwman, R.; Giron, I.H.; et al. Image quality in CT: From physical measurements to model observers. *Phys. Medica* **2015**, *31*, 823–843. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.