Journal of Computing and Security

Research Article, 2026, Volume 13, Number 2 (pp. 63–74)

http://www.jcomsec.org

Enhancing Low-Resolution Persian License Plates via Diffusion Models

Mahsham Kushki [a] Esmat Rashedi [a,*] Elham Shabaninia [b] Mehdi Kamandar [a]

[a] *Department of Telecommunications and Electronics Engineering, Faculty of Electrical and Computer Engineering, Graduate University of Advanced Technology, Kerman, Iran.*

[b] *Department of Applied Mathematics, Faculty of Modern Sciences and Technologies, Graduate University of Advanced Technology, Kerman, Iran.*

ARTICLE INFO.

Article history:
Received: 27 February 2026
Revised: 02 July 2026
Accepted: 27 July 2026
Published Online: 03 September 2026

Keywords:
Super-Resolution, Deep Learning, Diffusion Models, Persian License Plate Images.

ABSTRACT

With the rapid growth of urbanization and the increasing number of vehicles, automatic license plate recognition systems have become an essential component of traffic monitoring and management. However, the accuracy of these systems decreases when the captured plate images have low resolution or are affected by environmental noise. This study investigates the use of diffusion models to enhance the resolution of Iranian license plate images under such challenging conditions. In the proposed framework, a diffusion-based denoising process progressively reconstructs and refines image details iteratively. Experimental evaluations demonstrate that the proposed method outperforms well-known approaches such as Bicubic interpolation and deep learning-based super-resolution models, including SRGAN, Real-ESRGAN, BSRGAN, and SwinIR, in terms of perceptual quality and objective metrics. Specifically, the model achieves a PSNR of 31.2458 and an SSIM of 0.8621. Moreover, OCR results indicate that the reconstructed images produced by the diffusion model lead to better character readability and enhanced recognition accuracy.

1 Introduction

Nowadays, with the expansion of urbanization and the increase in the number of vehicles, traffic monitoring and transportation management have become critical issues in societies. This rapid growth requires innovative and efficient solutions in traffic monitoring, transportation system management, and control of traffic violations. In the present era, advanced computer vision techniques provide fast and accurate solutions for license plate recognition (LPR). This capability enhances tasks such as implementing intelligent parking systems or optimizing transportation systems [1, 2]. Moreover, it facilitates vehicle surveillance for enhancing security and managing traffic violations [1].

Several steps are involved in vehicle license plate recognition. First, pre-processing is applied to the image to enhance its quality and remove any noise. Next, the detection stage identifies the location of the license plate within the image. After that, the process moves to character recognition, where the characters on the plate are identified and extracted. Finally, post-processing is performed to correct any po-

* Corresponding author.
Email addresses: m.koushki@student.kgut.ac.ir (M. Kushki), e.rashedi@kgut.ac.ir (E. Rashedi), e.shabaninia@kgut.ac.ir (E. Shabaninia), m.kamandar@kgut.ac.ir (M. Kamandar)

https://dx.doi.org/10.22108/jcs.2026.148544.1198
ISSN: 2322-4460

tential errors and optimize the final results. During the pre-processing stage, image noise is reduced, and the resolution is enhanced. Increasing the resolution is considered one of the most important steps in the OCR process, as higher image quality has a direct impact on the accuracy of recognizing and correctly extracting license plate characters.

The low quality of input images is caused by poor lighting conditions, environmental pollution, and improper camera positioning. Additionally, the small size of the license plate in the image reduces details and clarity, making its detection and recognition difficult. Hence, in the vehicle license plate recognition process, image resolution plays a crucial role, as higher image resolution directly increases the accuracy of recognizing and extracting plate characters. Recently, significant advancements in deep learning and increased computational power have played a crucial role in improving image super-resolution processes. These methods employ models such as Convolutional Neural Networks (CNNs) to establish a comprehensive end-to-end mapping between low-resolution (LR) images and high-resolution (HR) images [2]. Generative Adversarial Networks (GANs), such as SRGAN [3] and ESRGAN [4], have significantly enhanced the quality of the reconstructed images.

Diffusion models (DMs), with their introduction, have brought about a major transformation in the field of image generation, including super-resolution (SR), and have challenged the dominant position of GANs [5]. The Denoising Diffusion Probabilistic Model (DDPM) [6] has recently gained widespread attention. This model is used in various areas such as image synthesis [5], super-resolution [7], image-to-image [8], and more.

In this paper, the application of diffusion models for enhancing the resolution of Iranian license plates (IR-LPR) [9] is examined under conditions where vehicle license plate recognition is challenged by low image resolution and environmental noise. Furthermore, the performance of the proposed method is evaluated through experiments and comparisons with other approaches such as Bicubic [10], SRGAN [3], Real-ESRGAN [11], BSRGAN [12], and SwinIR [13]. Additionally, considering that optical character recognition (OCR) heavily depends on image quality and resolution, the OCR performance before and after image enhancement is compared. These findings demonstrate the high capability of diffusion models in reconstructing high-quality images and enhancing the accuracy of license plate recognition. Furthermore, the results provide a new direction for future research in advancing traffic monitoring systems and automatic license plate recognition.

The structure of this paper is arranged as follows: Section 2 describes a summary of related works on image super-resolution and especially license plate super-resolution. Section 3 presents the diffusion models. Section 4 presents the experiments and their results, which are discussed from both quantitative and qualitative perspectives. Section 5 concludes the paper.

2 Related Works

In this section, previous studies are reviewed. First, early image enhancement methods are examined, and then techniques for enhancing vehicle license plate images are reviewed.

2.1 Image Super-Resolution Methods

Image super-resolution can be realized through various methods, which are broadly categorized into classical methods and deep learning-based methods. Classical methods for super-resolution encompass a variety of approaches, including edge-based methods [14, 15], statistical methods [16], patch-based methods [17], interpolation-based methods [10], and sparse representation-based techniques [18]. With the emergence of deep learning and the increase in computational power, this field has undergone a fundamental transformation. Convolutional neural networks were first introduced several decades ago as an innovative approach to processing image data [2]. The prominent methods in this field include SRCNN [2], FSRCNN [19], and ESPCNN [20]. Following the success of CNN-based approaches, GAN-based methods were introduced in 2017, consisting of two main components: the generator and the discriminator. Moreover, GAN-based methods, such as SRGAN and ESRGAN, have achieved remarkable advancements in enhancing perceptual quality. However, they still face challenges such as mode collapse, high computational cost, and difficulty in convergence [21]. Recently, diffusion-based models have gained significant attention in the field of super-resolution [22]. Different types of diffusion models, including Denoising Diffusion Probabilistic Models (DDPM) [6], Score-Based Generative Models (SGMs) [23], and Stochastic Differential Equations (SDEs) [24], define the fundamental principles and key structures of the diffusion process. These models have been successfully utilized to improve the quality of image reconstruction in various tasks.

Diffusion probabilistic models were first introduced by Ho et al. in 2020 [6]. The SR3 method is an innovative approach to image super-resolution

that leverages DDPMs for conditional image generation. This method performs super-resolution through a stochastic iterative denoising process [7]. In the SinSR method, a simple and fast approach for super-resolution image reconstruction is presented, requiring only a single inference step. This method produces high-quality outputs by learning from the teacher model and directly utilizing real images [25]. A method named ResShift has been proposed, which performs image super-resolution with high efficiency using only 15 sampling steps. In this method, instead of using Gaussian noise, the HR image is degraded toward the LR image, which significantly reduces the length of the diffusion process [26].

2.2 License Plate Image Super Resolution

Some studies examine super-resolution in license plate images. Enhancing the resolution of vehicle license plate images through image processing techniques has a substantial impact on achieving more accurate recognition and identification of license plates. Yang et al. [27] introduced the MSRCNN method, which employs a multi-scale convolutional neural network inspired by the Inception architecture of GoogLeNet to enhance the resolution of license plate images. In this model, convolutional layers with different kernel sizes capture fine and coarse image features at multiple spatial scales. The extracted features are then fused through non-linear activation functions such as ReLU to preserve both local textures and global structural information, resulting in higher-quality reconstructed images. Experimental results demonstrate that this approach significantly improves the clarity of low-resolution license plates and enhances recognition accuracy in surveillance systems. A new method is proposed for enhancing the resolution of license plate images using GANs and the pre-trained VGG-19 network. The proposed approach utilizes MSE and PSNR to reduce content loss while preventing excessive image smoothing, thereby preserving pixel data [28].

Mei et al. [29] proposed a GAN-based approach to enhance the quality of vehicle license plate images. In this method, the generator first synthesizes high-resolution images to learn detailed spatial and textural representations. These images are then downsampled to create realistic low- and high-resolution training pairs, enabling supervised learning of the reconstruction process. The generator subsequently reconstructs high-quality images from low-resolution inputs, while the discriminator evaluates the realism of the outputs and provides adversarial feedback to iteratively refine the generator's performance. The results indicated that this method enhances visual quality and OCR accuracy compared to interpolation methods, making it a valuable tool for security forces in extracting information from surveillance images.

Pan et al. [30] proposed a GAN-based method for enhancing the resolution of license plate images and improving the recognition rate of low-quality plates. They designed an n-stage random combination degradation model (n-RCD) to simulate the characteristics of low-resolution license plates in natural environments. Then, they developed a new network called LPSRGAN to optimize SRGAN and enhance image reconstruction. Additionally, a perceptual OCR loss function was introduced to preserve both character details and accuracy.

In a study conducted in Taiwan by Wang and colleagues, the low image quality of road surveillance cameras often hindered accurate license plate recognition. To address this issue, SR technology was utilized. This study evaluated the performance of three SR models: Real-ESRGAN, A-ESRGAN, and Star SRGAN in enhancing the resolution of license plate images and improving recognition accuracy, aiming to identify the most suitable model for this application [31]. Another method was proposed by AlHalawani and colleagues, focusing on enhancing the resolution of license plate images using SR technology and diffusion models. They utilized deep learning models to restore lost details in low-resolution license plate images. The results demonstrated that their approach improved license plate recognition accuracy and outperformed traditional techniques as well as other methods based on CNNs and GANs [32].

3 Diffusion-Based License Plate Super-Resolution

In this section, an approach is presented that has been specifically designed to enhance the resolution of Persian license plate images by using diffusion models (see Figure 1).

Assume that a paired dataset of high-resolution (HR) and corresponding low-resolution (LR) images is available, derived from the IR-LPR dataset specifically designed for Persian license plate images [9]. Here, x denotes the degraded low-resolution image, and z denotes its corresponding high-resolution version. The degradation typically results from adding noise to the HR image or reducing its resolution, as explained below [32].

$$x = G(z, \theta) \tag{1}$$

Additionally, θ represents the degradation parameters. In the super-resolution task, using a diffusion

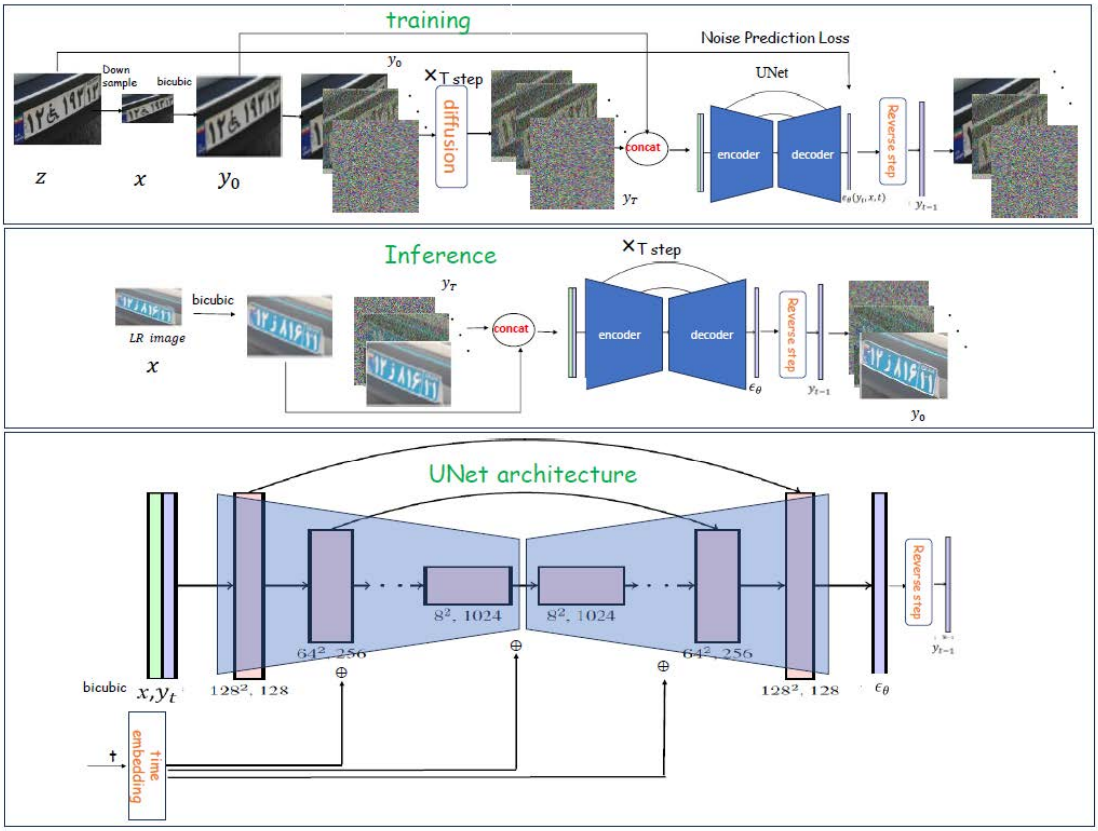


Figure 1. Overview of the SR framework. (Top) Training phase where the network ϵ_θ learns to predict the added noise. (Middle) Inference phase showing the iterative refinement. (Bottom) Detailed U-Net architecture [7]. The LR input x is bicubically upsampled and concatenated with y_t at each step. The network predicts the noise ϵ_θ, which is then used in the reverse diffusion step to compute y_{t-1}.

model, an attempt is made to reconstruct y_0, which is considered an approximation of the high-resolution image z from x. Although many existing methods focus on pixel-level details, this often leads to the neglect of the image's structural features. Diffusion models, based on non-equilibrium thermodynamics, consist of two phases: the forward noise application phase and the reverse noise removal phase.

Forward Diffusion: Starting from the initial input, the LR input image is enlarged to match the dimensions of the HR image using bicubic interpolation. Spherical Gaussian noise with variance β_t is added at each step of the Markov chain. At step t, y_t is generated by adding noise to y_{t-1} as follows [32]:

$$q\left(y_t \mid y_{t-1}\right) = \mathcal{N}\left(y_t; \mu_t = \sqrt{1-\beta_t}\, y_{t-1}, \Sigma_t = \beta_t I\right) \tag{2}$$

The Markov chain generates a step-by-step process that starts from the bicubic of the initial low-resolution image y_0 and progresses to the image y_T at the final time step. In this process, Gaussian noise is added at each step, gradually degrading the image and transferring it from the original space to a noisy space close to a Gaussian distribution. The goal of this chain is to produce a sequence of degraded images, where each image is generated using information from the previous step along with the added noise (as shown in Figure 2).

$$q\left(y_{1:T} \mid y_0\right) = \prod_{t=1}^{T} q\left(y_t \mid y_{t-1}\right) \tag{3}$$

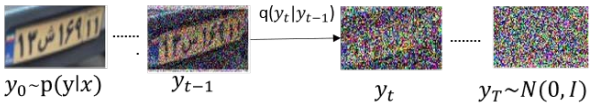


Figure 2. The sequential addition of Gaussian noise to the original image via a Markov chain, gradually transforming the low-resolution input into a distribution that approximates a Gaussian profile.

This chain generates a sequence from y_0 to y_T. Using the reparameterization trick and the notation

$\alpha_t = 1 - \beta_t$ and $\bar{\alpha}_t = \prod_{s=1}^{t} \alpha_s$, the following equations can be concisely defined [6].

$$\begin{aligned} y_t &= \sqrt{1-\beta_t}\, y_{t-1} + \sqrt{\beta_t}\, \epsilon_{t-1}, \qquad \epsilon_t \sim \mathcal{N}(0,1) \\ &= \sqrt{\alpha_t \alpha_{t-1}}\, y_{t-2} + \sqrt{\alpha_t(1-\alpha_{t-1})}\, \epsilon_{t-2} + \sqrt{1-\alpha_t}\, \epsilon_{t-1} \\ &= \dots \\ &= \sqrt{\bar{\alpha}_t}\, y_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon_0 \end{aligned} \tag{4}$$

In this section, by defining α_t and $\bar{\alpha}_t$, it becomes possible to determine y_t at each step based on β_t. Therefore, y_t is generated through the following process.

$$y_t \sim q\left(y_t \mid y_0\right) = \mathcal{N}\left(y_t; \sqrt{\bar{\alpha}_t}\, y_0, \left(1-\bar{\alpha}_t\right) I\right) \tag{5}$$

Reverse Diffusion: In the reverse, samples are generated from the original data distribution, which corresponds to the HR image in the super-resolution task. Starting from a noisy sample drawn from a Gaussian distribution, the model progressively removes noise through a sequence of denoising steps to reconstruct the original high-resolution image. This reverse process is modeled as a conditional Markov chain.

This operation, which is modeled as a conditional Markov chain under the model P_θ (implemented using a convolutional neural network), reconstructs the original data by progressively removing noise from the initial sample y_T, as illustrated in Figure 3.

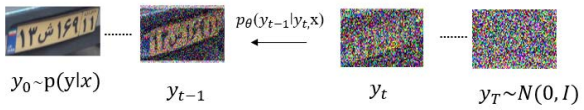


Figure 3. The reverse of the diffusion process. The model progressively removes noise using a network from a sample drawn from a Gaussian distribution, and ultimately reconstructs the original high-resolution image.

$$p_\theta\left(y_{0:T} \mid x\right) = p\left(y_T\right) \prod_{t=1}^{T} p_\theta\left(y_{t-1} \mid y_t, x\right) \tag{6}$$

where x represents the LR input image and y_T denotes the noisy HR sample at timestep t. The initial state y_T is sampled from a standard Gaussian distribution:

$$p\left(y_T\right) = \mathcal{N}\left(y_T; 0, I\right) \tag{7}$$

At each timestep, the reverse transition is modeled as a Gaussian distribution:

$$p_\theta\left(y_{t-1} \mid y_t, x\right) = \mathcal{N}\left(y_{t-1}; \mu_\theta\left(y_t, x, t\right), \beta_t I\right) \tag{8}$$

where $\mu_\theta(y_t, x, t)$ represents the predicted mean and β_t is the variance associated with the diffusion schedule.

For the generative model to reconstruct the high-resolution image, it must have access to information from the original low-resolution input. Therefore, the reverse diffusion process is conditioned on the LR image x at each timestep. Instead of directly predicting the denoised image, the neural network is trained to estimate the noise added during the forward diffusion process. The network is defined as

$$\varepsilon_\theta\left(y_t, x, t\right) \tag{9}$$

which predicts the noise component ε presented in y_t. Following the formulation in diffusion-based super-resolution models such as SR3 [7], the mean of the reverse Gaussian distribution can be derived from the predicted noise:

$$\tilde{\mu}_\theta\left(y_t, x, t\right) = \frac{1}{\sqrt{\alpha_t}} \left(y_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}_t}}\, \varepsilon_\theta\left(y_t, x, t\right)\right) \tag{10}$$

The U-Net used in the diffusion model bears a strong resemblance to the structure employed in [7], specifically adapted for the image super-resolution task. In the contracting path, multi-scale features are progressively extracted from the LR input through successive convolutional and down-sampling layers. This pathway captures both fine local textures and global structural information essential for accurate image reconstruction. The expanding path performs progressive up-sampling and feature fusion via skip connections, integrating detailed spatial information from early layers with deeper semantic representations. The upsampled LR image is concatenated with the noisy sample y_t at each diffusion step and provided as input to the U-Net together with the timestep embedding. The network outputs the predicted noise $\varepsilon_\theta(y_t, x, t)$. The training objective is defined as the mean squared error between the true noise and the predicted noise:

$$\mathcal{L}(\theta) = \mathbb{E}_{y_0, \epsilon_t, t}\left[\left\|\epsilon_t - \varepsilon_\theta\left(y_t, x, t\right)\right\|_2^2\right] \tag{11}$$

By minimizing this objective, the model learns to accurately estimate the noise at each diffusion step. This enables the reverse diffusion process to iteratively remove the noise and reconstruct a high-

resolution license plate image conditioned on the corresponding low-resolution input.

4 Experiments

In this section, the dataset is introduced, then the evaluation metrics are described for image super-resolution. Next, various methods are compared to assess how much the license plate image quality improved. Finally, to examine the impact of these enhancements on OCR results, the OCR outputs are compared before and after image resolution enhancement.

4.1 Dataset

The IR-LPR [9] dataset is gathered over the period from 2018 to 2022. This dataset includes 19,937 vehicle images with fully annotated license plates and 27,745 license plate characters. These data can be utilized for various applications, such as vehicle identification, automated toll collection, stolen vehicle detection, traffic law enforcement, and access control to private areas. Iranian license plates follow a standard format and consist of two sections. The first section is a unique combination of two digits on the left, an alphabetical character in the middle, and three digits on the right. Additionally, the background color of the plates varies based on the type and usage of the vehicle; for example, private cars have a white background, while taxis and public transportation vehicles have a yellow background. The dataset covers two different lighting conditions, including 16,845 images captured during the day and 4,122 images captured at night. Additionally, a separate dataset containing 27,745 license plate images has been prepared to enhance the license plate character recognition process. We randomly select 2000 samples as test data. The original images are downsampled by a factor of 8, resulting in images with dimensions of $64 \times 64 \times 3$. These downsampled images are used as inputs to the diffusion model. Figure 4 presents some sample images from the dataset.

Figure 4. Sample Images from the IR-LPR Dataset.

4.2 Evaluation Metrics

Image quality is a multidimensional concept that encompasses features such as sharpness, contrast, and noise reduction [33]. Therefore, an accurate evaluation of SR models based on the quality of the generated images is a complex and challenging process. To accurately and comprehensively evaluate the performance of the model, we used the following metrics:

Peak Signal-to-Noise Ratio: Peak Signal-to-Noise Ratio (PSNR) is one of the most commonly used metrics for evaluating image quality and reconstruction. This metric defines the ratio of the maximum pixel value L to the mean squared error (MSE) between the SR image y and the HR image z [33], where MSE is the sum of the squared differences between each pair of pixels in the two images divided by N (where N is the total number of pixels). This metric focuses on pixel differences, which often do not align with human perceptual quality, as even the smallest pixel shifts can reduce the PSNR value without affecting the quality perceived by the human eye. It can be expressed as follows:

$$\mathrm{PSNR}(z,y) = 10 \cdot \log_{10}\left(\frac{L^2}{\frac{1}{N}\sum_{i=1}^{N}[z-y]^2}\right) \quad (12)$$

SSIM Index: Structural Similarity (SSIM) is a commonly used method for image evaluation that focuses on the differences in structural features between images. This metric independently calculates SSIM by comparing the luminance, contrast, and structural comparison measures of the image [33].

$$\mathrm{SSIM}(z,y) = [l(z,y)]^{\alpha} \cdot [c(z,y)]^{\beta} \cdot [s(z,y)]^{\gamma} \quad (13)$$

Here α, β, and γ, each greater than zero, are parameters that can be adjusted to control the relative importance of the components. This function uses the three components of luminance l, contrast c, and structure s comparison measures.

MS-SSIM: It is an image quality assessment metric that operates on multiple scales. By analyzing image quality at different resolution levels, it provides higher accuracy and is used for evaluating super-resolution methods and image compression. MS-SSIM can be written as follows [34].

$$\text{MS-SSIM}(z,y) = l_M(z,y)^{\alpha_M} \cdot \prod_{j=1}^{M}[c_j(z,y)]^{\beta_j}[s_j(z,y)]^{\gamma_j} \quad (14)$$

In this equation, l_M denotes the luminance comparison at scale M. c_j and s_j describe the image's

contrast and structure comparisons at scale j, respectively. The constants α_M, β_j, and γ_j are used to specify the contribution of each of these components to the final result.

4.3 Training Details

In this section, the model settings and details related to model training are discussed. The LP-LPR dataset is used for training the model. This dataset contains diverse images taken from different angles. The model is implemented using PyTorch. Additionally, to meet the computational demands of the diffusion model and comparison models, Google Colab, Google Colab Pro, and the NVIDIA Quadro RTX graphics card are used.

For the diffusion model, the image size is initially reduced from $512 \times 512 \times 3$ to $64 \times 64 \times 3$. The model training process is carried out over 910,000 iterations, based on Saharia et al. [7]. Table 1 presents the main architectural details.

Table 1. Main Architectural Details.

Parameter	Quantity
Learning rate	3e-6
Batch size	2
# workers	8
# Channels	3
Low-resolution size	64
High-resolution size	512
Optimizer	Adam

4.4 Visualization of the Reconstruction Process

In Figure 5, there is an input image with dimensions $64 \times 64 \times 3$, shown in (a). Then, using the denoising process in the diffusion model, as illustrated in (b), the image is processed. This process results in a high-resolution image with dimensions $512 \times 512 \times 3$, shown in (c). By comparing this image with the Ground Truth (d) image, it is observed that they are visually very similar, and the differences between them are hardly noticeable.

4.5 Results

In the following, the quantitative analysis is presented first, followed by the qualitative analysis of the results for test images. The results are reported for test data.

Quantitative Comparison: This section compares our results with the results obtained from four selected methods, including Bicubic [10], SRGAN [3], Real-ESRGAN [11], BSRGAN [12], and SwinIR [13]. The results are presented in Table 2.

Table 2. The quantitative performance of different methods for image super-resolution.

Method	PSNR	SSIM	MS-SSIM
Bicubic	22.1678	0.7275	0.8878
SRGAN	29.4276	0.8150	0.8712
Real-ESRGAN	25.1024	0.7513	0.8391
BSRGAN	25.9881	0.7833	0.8655
SwinIR	21.5635	0.7175	0.8132
Diffusion Model	**31.2458**	**0.8621**	**0.9023**

For evaluating the results, the PSNR, SSIM, and MS-SSIM metrics are used, as detailed in Section 4. The results demonstrate that the diffusion model outperforms the compared methods. Specifically, the PSNR value for the upsampling model is improved by 9.078 compared to the Bicubic method, by 1.8182 compared to the SRGAN method, by 6.1434 compared to the Real-ESRGAN method, by 5.2577 compared to the BSRGAN, and by 9.6823 compared to the SwinIR method. In addition, the SSIM metric of the diffusion model outperforms the Bicubic method by 0.1346, the SRGAN model by 0.0471, the Real-ESRGAN model by 0.1108, the BSRGAN model by 0.0788, and the SwinIR model by 0.1446. The diffusion model performs better and more efficiently than traditional and advanced methods in generating high-resolution images, thereby improving the quality of license plate images.

On the other hand, the MS-SSIM metric outperforms the Bicubic method by 0.0145, the SRGAN method by 0.0311, the Real-ESRGAN method by 0.0632, the BSRGAN method by 0.0368, and the SwinIR method by 0.0891. Bicubic is a classical upsampling technique that estimates pixel values based on neighboring pixels; however, it cannot recover lost high-frequency information, resulting in blurred edges and limited detail restoration. SRGAN improves perceptual image quality using adversarial learning and perceptual loss, producing sharper images than interpolation-based methods, but it may introduce artificial textures and visual artifacts. Real-ESRGAN is designed for blind super-resolution in real-world images and uses high-order degradation modeling; although it enhances perceptual sharpness, it may still generate unrealistic textures in some regions. Similarly, BSRGAN simulates complex

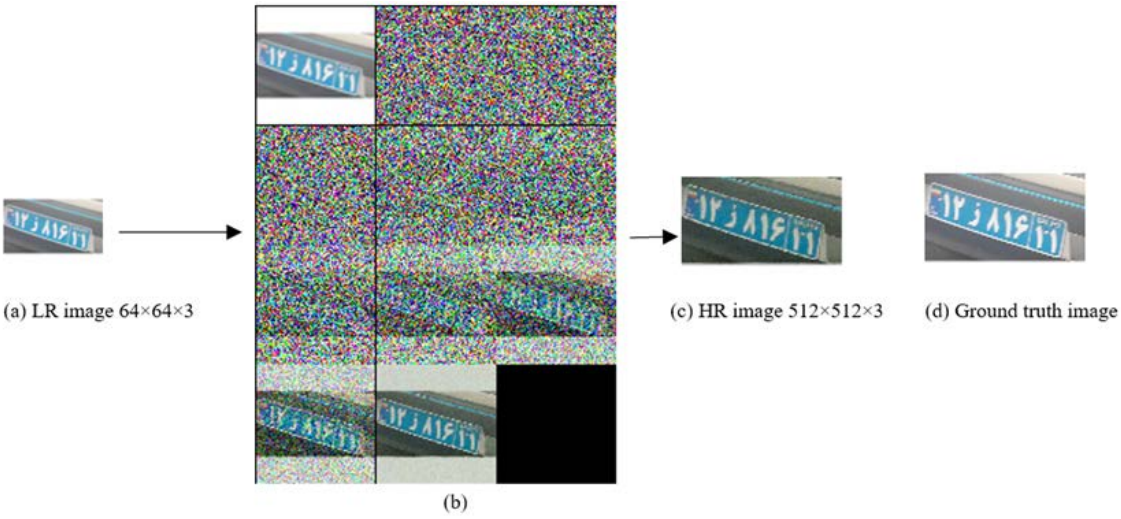


Figure 5. The various stages of the image reconstruction process. (a) Low-resolution image ($64 \times 64 \times 3$), (b) denoising process via the diffusion model, (c) reconstructed high-resolution image ($512 \times 512 \times 3$), and (d) ground truth image for comparison.

degradation processes to handle real-world degraded images, but it may face challenges in consistently preserving structural details. SwinIR, which is based on a transformer architecture, performs image restoration using Swin Transformer blocks; however, it may have limitations in the stable refinement of local structures. In contrast, the diffusion-based framework reconstructs images through an iterative denoising process, in which structural details and edges are gradually refined. This leads to more stable reconstructions, fewer artifacts, and better preservation of fine details compared with the examined methods.

Overall, the evaluation results indicate that the diffusion model achieves acceptable performance in the quantitative metrics PSNR, SSIM, and MS-SSIM compared with Bicubic, SRGAN, Real-ESRGAN, BSRGAN, and SwinIR. Improvements in these metrics indicate better reconstruction of image details, more effective preservation of structural information, and enhanced perceptual image quality. This performance can be attributed to the diffusion model's ability to gradually learn image details, reconstruct lost information step by step, and preserve structural and textural consistency during the generation process. By leveraging a progressive denoising process, the model improves the reconstruction of fine details and the overall image structure, providing satisfactory results compared with both traditional methods and advanced deep learning-based approaches.

Qualitative Results: The results presented in Figure 6 show the visual outputs of various methods for some sample images. After reviewing the images produced by other methods, it is observed that the proposed method produces fair results. Diffusion-based frameworks rely on an iterative denoising mechanism, which enables a progressive and stable image reconstruction process rather than a single-pass inference. In contrast to GAN-based architectures, which are often susceptible to training instability and the generation of artifact-prone, hallucinated textures, diffusion approaches benefit from a more robust optimization landscape. This methodology effectively mitigates unwanted artifacts, thereby supporting the structural integrity of character boundaries and high-frequency details. Furthermore, compared to Transformer-based approaches, diffusion models facilitate more effective local feature refinement through their multi-step reconstruction process, which contributes to better preservation of sharp edges and fine structural elements that are essential for OCR readability.

Finally, the quantitative and qualitative evaluations show that the images generated by the diffusion model accurately preserve all the visual and content-related features of the original image. This indicates that the diffusion model has the capability to generate high-resolution images in a way that preserves the intrinsic and structural features of the original image without any alterations or loss of accuracy during the reconstruction process.

Computational Complexity: Table 3 illustrates the computational complexity of different methods. In terms of computational cost, the compared approaches exhibit notable differences. The diffusion model requires the highest amount of computation and has a longer execution time compared to other methods. This high complexity is inherent to the structure of diffusion-based models, as the image reconstruction process is performed iteratively and step-by-step, with each stage involving a full forward

Figure 6. The Visual Outputs of Various Methods for Some Sample Images.

pass through the network. Consequently, although diffusion models are inherently computationally heavy and time-consuming, this cost leads to significant improvements in reconstruction quality and the preservation of fine image details.

Table 3. The Parameters and Total Inference Time for an Image.

	Bicubic	SRGAN	Real-ESRGAN	BSRGAN	SwinIR	Diffusion Model
Parameters	0	16.85M	16.73M	16.7M	11.8M	628M
Inference Time (s)	0.20	0.25	0.24	0.24	0.22	0.83

OCR: As previously stated, the accuracy of license plate recognition systems is highly dependent on the quality of the input image [35]. In this section, a comparison is presented between the vehicle license plate recognition results before and after using image super-resolution. OCR is a process that converts document images into searchable and editable text; for this reason, this technology is used as an efficient tool for processing scanned documents [36].

To evaluate the impact of the proposed super-resolution method on downstream recognition, license plate OCR results before and after image enhancement are compared. For this purpose, the OCR model from [37] is used, which was trained on the IR-LPR dataset and follows a hybrid vision-and-language-model framework. In this approach, the visual recognition stage extracts the initial plate characters, and a language-model-based post-processing module then corrects errors according to the structural patterns of Iranian license plates. Applying the same OCR pipeline to both the original low-resolution images and the super-resolved outputs allows us to assess how improvements in visual quality contribute to more reliable recognition. Figure 7 summarizes these results.

As shown in Figure 7, a comprehensive comparison has been presented between the vehicle license plate recognition results before and after applying the image super-resolution technique. The test data OCR results are compared to human-provided labels, resulting in an accuracy of 86.8 before image super-resolution. After resolution enhancement, an accuracy of 90.5 is achieved. In this study, the direct impact of improved image clarity on input quality and consequently on the performance of the license plate recognition system was investigated. The primary objective of this comparison is to evaluate how enhancing image clarity can increase both the accuracy and efficiency of license plate recognition. The obtained results demonstrate that improved image clarity leads to higher recognition accuracy and enhanced system performance, with significant differences observed between the results before and after applying this technique. These findings indicate that by elevating the quality of input images, the system can extract more detailed and precise information from the license plates, resulting in overall better performance.

5 Conclusions

In this paper, a deep model is applied that leverages diffusion techniques to reconstruct high-quality license plate images from noisy and degraded inputs. Our approach involves a carefully designed diffusion process that iteratively refines the input image, reducing noise and restoring critical details, which is essential for effective license plate recognition.

The results indicate that the diffusion model outperforms other methods such as Bicubic, SRGAN, Real-ESRGAN, BSRGAN, and SwinIR. The enhanced images produced not only display increased clarity and detail, but also significantly contribute to improving the accuracy and efficiency of subsequent vehicle license plate recognition systems. This improvement is especially important in real-world conditions where a decline in image quality can seriously affect recognition performance.

Despite the promising results, one of the notable drawbacks of this approach is its high computational cost. Although the diffusion process is efficient, it re-

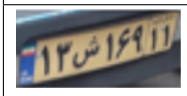
Original image	Before SR		After SR		
	OCR result	# error	SR images	OCR result	# error
	**12 ی 169 11**	**2**		**13 ش 169 11**	**0**
	**22 الف 117 70**	**4**		**14 الف 117 79**	**0**
	**18 ت 184 11**	**1**		**18 ث 181 11**	**0**

Figure 7. A comparison of the vehicle license plate recognition results before and after image super-resolution.

Table 4. A comparison of the vehicle license plate recognition OCR results of test images after different image super-resolution methods.

Bicubic	SRGAN	Real-ESRGAN	BSRGAN	SwinIR	Diffusion model
87.7	89.5	88.02	88.5	87.4	90.5

quires significant processing power and time, which may limit its applicability in real-time or resource-constrained environments. Future research will focus on optimizing this process, exploring techniques such as model compression, accelerated algorithms, or hybrid approaches to reduce the computational demands without compromising the quality of the reconstructed images.

References

[1] P. Sikora, L. Malina, M. Kiac, Z. Martinasek, K. Riha, J. Prinosil, L. Jirik, and G. Srivastava. Artificial intelligence-based surveillance system for railway crossing traffic. *IEEE Sensors Journal*, 21(14):15515–15526, 2021. doi:10.1109/JSEN.2020.3031861.

[2] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2016. doi:10.1109/TPAMI.2015.2439281.

[3] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4681–4690, 2017. doi:10.1109/CVPR.2017.19.

[4] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *Computer Vision – ECCV 2018 Workshops*, volume 11133, pages 63–79. Springer, 2019. doi:10.1007/978-3-030-11021-5_5.

[5] P. Dhariwal and A. Nichol. Diffusion models beat GANs on image synthesis. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pages 8780–8794, 2021. doi:10.48550/arXiv.2105.05233.

[6] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 33, pages 6840–6851, 2020. doi:10.48550/arXiv.2006.11239.

[7] C. Saharia, J. Ho, W. Chan, T. Salimans, D. J. Fleet, and M. Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(4):4713–4726, 2023. doi:10.1109/TPAMI.2022.3204461.

[8] C. Saharia, W. Chan, H. Chang, C. Lee, J. Ho, T. Salimans, D. J. Fleet, and M. Norouzi. Palette: Image-to-image diffusion models. In *ACM SIGGRAPH 2022 Conference Proceedings*, pages 1–10, 2022. doi:10.1145/3528233.3530757.

[9] M. Rahmani, M. Sabaghian, S. M. Moghadami, M. M. Talaie, M. Naghibi, and M. A. Keyvanrad. IR-LPR: A large scale Iranian license plate recognition dataset. In *2022 12th International Conference on Computer and Knowledge Engineering (ICCKE)*, pages 53–58. IEEE, 2022. doi:10.1109/ICCKE57176.2022.9960129.

[10] R. Keys. Cubic convolution interpolation for digital image processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6):1153–1160, 1981. doi:10.1109/TASSP.1981.1163711.

[11] X. Wang, L. Xie, C. Dong, and Y. Shan.

Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1905–1914, 2021. doi:10.1109/ICCVW54120.2021.00217.

[12] K. Zhang, J. Liang, L. Van Gool, and R. Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4791–4800, 2021. doi:10.1109/ICCV48922.2021.00475.

[13] J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte. SwinIR: Image restoration using Swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1833–1844, 2021. doi:10.1109/ICCVW54120.2021.00210.

[14] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE, 2008. doi:10.1109/CVPR.2008.4587659.

[15] G. Freedman and R. Fattal. Image and video upscaling from local self-examples. *ACM Transactions on Graphics (TOG)*, 30(2):12:1–12:11, 2011. doi:10.1145/1944846.1944852.

[16] K. I. Kim and Y. Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1127–1133, 2010. doi:10.1109/TPAMI.2010.25.

[17] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 275–282. IEEE, 2004. doi:10.1109/CVPR.2004.1315043.

[18] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing*, 19(11):2861–2873, 2010. doi:10.1109/TIP.2010.2050625.

[19] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *Computer Vision – ECCV 2016*, pages 391–407. Springer, 2016. doi:10.1007/978-3-319-46475-6_25.

[20] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang. 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 (CVPR)*, pages 1874–1883, 2016. doi:10.1109/CVPR.2016.207.

[21] S. Frolov, T. Hinz, F. Raue, J. Hees, and A. Dengel. Adversarial text-to-image synthesis: A review. *Neural Networks*, 144:187–209, 2021. doi:10.1016/j.neunet.2021.07.019.

[22] B. B. Moser, A. S. Shanbhag, F. Raue, S. Frolov, S. Palacio, and A. Dengel. Diffusion models, image super-resolution, and everything: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 36(7):11793–11813, 2025. doi:10.1109/TNNLS.2024.3476671.

[23] Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, pages 11918–11930, 2019. doi:10.48550/arXiv.1907.05600.

[24] Y. Song, J. Sohl-Dickstein, D. P. Kingma, A. Kumar, S. Ermon, and B. Poole. Score-based generative modeling through stochastic differential equations. In *International Conference on Learning Representations (ICLR)*, 2021. doi:10.48550/arXiv.2011.13456.

[25] Y. Wang, W. Yang, X. Chen, Y. Wang, L. Guo, L.-P. Chau, Z. Liu, Y. Qiao, A. C. Kot, and B. Wen. SinSR: Diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25796–25805, 2024. doi:10.1109/CVPR52733.2024.02437.

[26] Z. Yue, J. Wang, and C. C. Loy. ResShift: Efficient diffusion model for image super-resolution by residual shifting. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pages 13294–13307, 2023. doi:10.48550/arXiv.2307.12348.

[27] Y. Yang, P. Bi, and Y. Liu. License plate image super-resolution based on convolutional neural network. In *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pages 723–727. IEEE, 2018. doi:10.1109/ICIVC.2018.8492768.

[28] B. K. Shah, A. Yadav, and A. K. Dixit. License plate image super-resolution using generative adversarial network (GAN). In *2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, pages 1139–1143. IEEE, 2022. doi:10.1109/ICAAIC53929.2022.9792759.

[29] Y. Mei, M. Moelter, and R. J. Haddad. License plate image resolution enhancement using super-resolution generative adversarial network. In *IEEE SoutheastCon 2024*, pages 1262–1267. IEEE, 2024. doi:10.1109/SoutheastCon52093.2024.10500226.

[30] Y. Pan, J. Tang, and T. Tjahjadi. LP-SRGAN: Generative adversarial networks

for super-resolution of license plate image. *Neurocomputing*, 580:127426, 2024. doi:10.1016/j.neucom.2024.127426.

[31] C.-H. Wang. Using super-resolution imaging for recognition of low-resolution blurred license plates: A comparative study of Real-ESRGAN, A-ESRGAN, and StarSR-GAN. *arXiv preprint arXiv:2403.15466*, 2024. doi:10.48550/arXiv.2403.15466.

[32] S. AlHalawani, B. Benjdira, A. Ammar, A. Koubaa, and A. M. Ali. DiffPlate: A diffusion model for super-resolution of license plate images. *Electronics*, 13(13):2670, 2024. doi:10.3390/electronics13132670.

[33] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi:10.1109/TIP.2003.819861.

[34] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, volume 2, pages 1398–1402. IEEE, 2003. doi:10.1109/ACSSC.2003.1292216.

[35] A. Afkari-Fahandari, E. Shabaninia, F. Asadi-Zeydabadi, and H. Nezamabadi-Pour. A comprehensive survey of transformers in text recognition: Techniques, challenges, and future directions. *ACM Computing Surveys*, 58(5):111:1–111:42, 2026. doi:10.1145/3771273.

[36] F. Asadi-Zeydabadi, A. Afkari-Fahandari, A. Faraji, E. Shabaninia, and H. Nezamabadi-Pour. IDPL-PFOD2: A new large-scale dataset for printed Farsi optical character recognition. *arXiv preprint arXiv:2312.01177*, 2023. doi:10.48550/arXiv.2312.01177.

[37] F. Asadi-Zeydabadi, E. Shabaninia, and H. Nezamabadi-Pour. Enhancing license plate recognition using a language model-based approach. *Journal of Machine Vision and Image Processing*, 11(4):15–26, 2025. doi:10.52547/jmvip.11.4.15.

Mahsham Kushki received her B.Sc. degree in 2019 and her M.Sc. degree in System Communications in 2023 from the Graduate University of Advanced Technology (HITEC), Kerman, Iran. She is currently pursuing a Ph.D. degree in System Communications at the same university. Her research interests include image processing, machine learning, and artificial intelligence.

Esmat Rashedi received her Ph.D. degree in Communication Engineering. She is currently an Associate Professor in the Faculty of Electrical and Computer Engineering, Graduate University of Advanced Technology in Kerman, Iran. Her research interests include image and video processing, pattern recognition, heuristic optimization algorithms, and deep neural networks.

Elham Shabaninia holds an M.Sc. degree in Computer Engineering from Sharif University of Technology (2009) and a Ph.D. degree in Artificial Intelligence from the University of Isfahan (2018). Following her postdoctoral research at Shahid Bahonar University of Kerman, she joined the Graduate University of Advanced Technology (HITEC) in Kerman in 2023, where she currently serves as an Assistant Professor in the Department of Applied Mathematics. Her research interests primarily focus on artificial intelligence, machine learning, and image processing.

Mehdi Kamandar received his Ph.D. degree in Electrical Engineering, specializing in System Communications. He is currently an Assistant Professor in the Department of Electrical Engineering at the Graduate University of Advanced Technology (HITEC) in Kerman, Iran. His research interests include biomedical signal processing, machine learning, and deep neural networks.