Emirates Journal for Engineering Research

Volume 29 | Issue 3 Article 2

8-18-2024

GA-GESRGAN: Document Images Super Resolution using Gabor Filters, ESRGAN Models and Genetic Algorithms

Zakia KEZZOULA
Universite M'hamed Bouguerra de Boumerdes, z.kezzoula@univ-boumerdes.dz

Djamel GACEB
Universite M'hamed Bouguerra de Boumerdes, d.gaceb@univ-boumerdes.dz

Ayoub TITOUN
F2i Institute, School of Computer Science, Digital & Commerce, Vincennes, France, ia_titoun@esi.dz

GA-GESRGAN: Document Images Super Resolution using Gabor Filters, ESRGAN Models and Genetic Algorithms

Cover Page Footnote

The authors extend their heartfelt thanks to LIMOSE Laboratory in the Computer Science Department at the University of Boumerdes for providing the opportunity to work on this paper

GA-GESRGAN: DOCUMENT IMAGES SUPER-RESOLUTION USING GABOR FILTERS, ESRGAN MODELS, AND GENETIC ALGORITHMS

KEZZOULA Zakia[1*], GACEB Djamel[1] and Titoun Ayoub[2]

[1]LIMOSE Laboratory, University M'Hamed Bougara of Boumerdes, ALGERIA

z.kezzoula@univ-boumerdes.dz, d.gaceb@univ-boumerdes.dz
[2]F2i Institute, School of Computer Science, Digital & Commerce, Vincennes, France
ia_titoun@esi.dz

(Received 19th Javier 2024 and Accepted on 18th August 2024)

GA-GESRGAN: تحسين الدقة الفائقة للصور الوثائقية باستخدام مرشحات جابور، نماذج ESRGANوالخوارزميات الجينية

ملخص

خلال العقد الماضي، شهدنا تقدمًا كبيرًا في تقنية تحسين الدقة الفائقة للصور، وذلك بفضل ظهور وتطوير نماذج التعلم العميق التي يمكنها التكيف مع تعقيد المهام وتحسين جودة الصور. يقدم هذا المقال نهجًا جديدًا لتحسين الدقة الفائقة للصور باسم GA-GESRGANلتعزيز جودة صور الوثائق من خلال منهجية متعددة الخطوات. في البداية، تتم معالجة صور الوثائق باستخدام مرشحات جابور لاستخراج الميزات عبر الترددات المكانية المختلفة. يتم بعد ذلك استخدام هذه الميزات المستخرجة لتدريب نماذج الشبكة التوليدية المعادية لتحسين الدقة الفائقة (ESRGAN). بعد الحصول على الصور المحسنة من خلال نماذجESRGAN ، يتم استخدام خوارزمية جينية لدمج هذه النتائج بشكل فعال. تسلط هذه المنهجية المبتكرة الضوء على التكامل بين تقنيات معالجة الصور التقليدية، نماذج التعلم العميق، وخوارزميات التحسين المتقدمة، مما يؤدي في النهاية إلى تحسينات كبيرة في جودة صور الوثائق. وتظهر النتائج الإمكانيات الكبيرة لهذا النهج.

Abstract

In the last decade, we have witnessed significant progress in image super-resolution, thanks in particular to the emergence and improvement of deep learning models, which can adapt to the complexity of tasks and improve image quality. This article presents a novel approach to image super-resolution GA-GESRGAN to enhance the quality of document images through a multi-step methodology. Initially, document images are processed using Gabor filters to extract features across various spatial frequencies. These extracted features are then utilized to train Enhanced Super-Resolution Generative Adversarial Network (ESRGAN) models. Once improved images are obtained through ESRGAN models, a Genetic Algorithm combines these results effectively. This innovative methodology highlights the synergy between traditional image processing techniques, deep learning models, and advanced optimization algorithms, ultimately leading to significant improvements in document image quality. The results demonstrate the potential of this approach.

Keywords: Document Images, Super Resolution, Gabor Filters, ESRGAN Models, Genetic Algorithms.

1. INTRODUCTION

Improving image resolution is a constant challenge in the field of computer vision. In many applications, including the preservation of historical archives, medical radiology, and security monitoring (see Figure 1), there exists a critical need to restore low-resolution and degraded images to sharp, detailed versions.

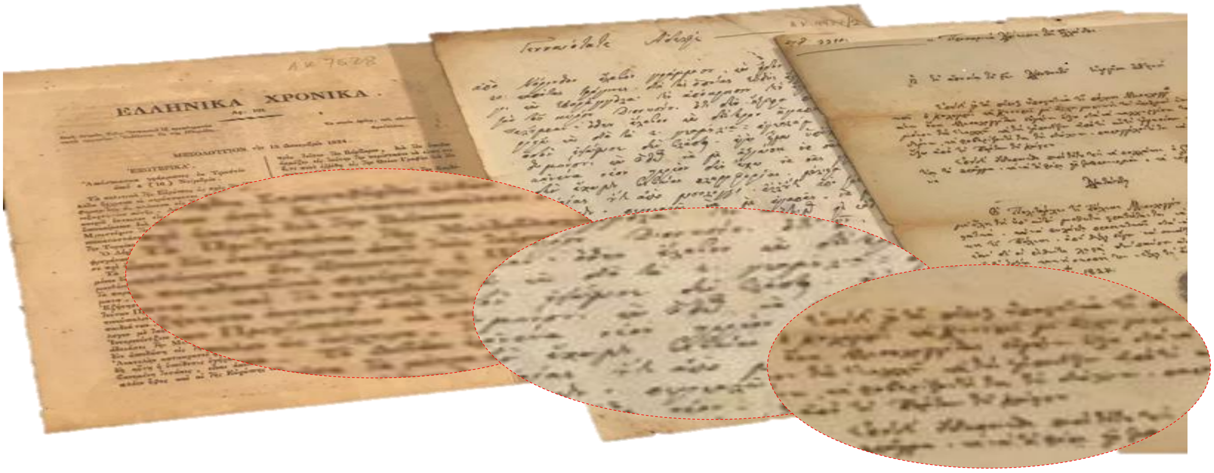

Figure 1 Low-resolution images examples.

This article presents a novel super resolution SR methodology that combines two bio-inspired approaches, deep learning, and genetic algorithm optimization to achieve this goal. The process starts with the use of Gabor filters based on mimicking how human neurons detect contours and textures in visual stimulus. These filters are applied to images to extract features at different spatial frequencies, including vertical, horizontal, and diagonals frequencies. This pre-processing step plays a crucial role in capturing the information essential for high-resolution image reconstruction.

These features are then used as input data for an ESRGAN model, which demonstrates exceptional competence in learning complex structures and capturing fine detail from large training datasets, marking a significant advance in super-resolution methodologies.

However, our methodology does not stop there. We employ a genetic algorithm to maximize the quality of the ESRGAN-enhanced photos once we have them. The genetic algorithm seeks to optimize the image quality metric known as PSNR (Peak Signal-to-Noise Ratio). To provide the sharpest, clearest images possible, it modifies the coefficients of the images produced by ESRGAN.

The method differs in that it solves the complex challenge of document image super-resolution by creatively combining deep learning, genetic algorithm optimization, and the Gabor filter. The results show how well it works to produce images of excellent quality, which is crucial for many real-world uses.

The next sections of the article will review existing super-resolution methods and give a detailed explanation of our approach to document image super-resolution. We will also present the results of our studies and the enhanced quality of document photos obtained using our methods.

2. STATE-OF-THE-ART SUPER RESOLUTION TECHNIQUES

Existing super-resolution approaches can be categorized into several groups based on their underlying techniques and methodologies. Here are the finest categories of super-resolution methods:

Multiple Input Images super-resolution refers to a class of image enhancement techniques that take multiple low-resolution images as input to generate a single high-resolution image as output. This approach is often used when several low-resolution images of the same scene or object are available, such as a burst of images from a camera or frames from a video.

Single Input Image: This category deals with techniques that use only one input image to produce a high-resolution output. Within this category:

- Interpolation techniques.
- Image transformation-based methods.
- Learning-based methods.

Image interpolation techniques are methods for estimating pixel values at non-integer coordinates in an image. There are two main categories [1]: adaptive and non-adaptive. Each of these methods has advantages and disadvantages in terms of image quality, blur, and processing speed.

Non-Adaptive interpolation: Non-adaptive algorithms use a fixed interpolation method, regardless of the nature of the image. This makes them faster in terms of processing, but can lead to a loss of detail in complex areas. These methods are often simpler to implement and may be more appropriate for real-time applications where speed is crucial. Several methods are available, including nearest neighbor interpolation [2], bilinear [3], bicubic [4], B-spline [5], Lanczos [6], Mitchell-Netravali [7], spline [8], Lagrange polynomial [9], natural neighbors [10] …etc. Each method has its advantages and disadvantages in terms of quality and complexity.

Adaptive interpolation: Adaptive algorithms adjust their interpolation method according to the characteristics of the image. For example, in areas with strong variations in color or texture, a more complex interpolation method can be used to preserve detail. On the other hand, in more uniform areas, simpler interpolation can be applied to reduce noise. The advantages of this approach are better image quality and reduced blurring, but it can be more computationally intensive. Adaptive interpolation methods have been widely explored in the literature to improve image quality, including approaches like Adaptive Windowed Sinc Filter for Image Interpolation [11], Fast super-resolution using an adaptive Wiener filter with robustness to local motion [12], New edge-directed interpolation[13].

Reconstruction-based super-resolution methods, also known as transformation methods, are approaches that use mathematical operations, such as convolution masks, to estimate and reconstruct a high-resolution image from a low-resolution image. Many algorithms have been proposed. We cite a few reference works in this category of techniques, such as those by Naik and Patel [14], Kumar et al. [15], Anbarjafari and, Demirel [16].

Methods of super-resolution that utilize learning algorithms have significantly enhanced the quality of image enlargement outcomes over traditional techniques. .These methods use deep neural networks to learn from training data and generate high-quality HR images. In the field of image super resolution, there has been a notable paradigm change over time from machine learning techniques to deep learning-based methods.

Deep learning has become a powerful tool for image super-resolution compared to machine learning-based approaches. This emerges from its capacity to learn and represent complicated image characteristics automatically, to adjust with different image formats, and to produce better results than traditional learning-based approaches. The development of Convolutional Neural Networks (CNNs) particularly designed for super-resolution has been one of the most significant advancements. The introduction of a CNN in the algorithm of Dong et al. in [17] marked a turning point in the field. CNNs facilitated the learning of complex techniques from large datasets, leading to high-quality super resolution results. The CNN-based super-resolution process begins by assembling a data set consisting of pairs of low and high-resolution images. These pairs are then used as a training set for the network. The architecture of the CNN model is critical. Network architectures, such as the VGG network [18], are specifically designed to extract features from images, using deconvolution layers to spatially extend the extracted features. Residual layers, such as those used in ResNet networks [19], can also be incorporated to increase the stability of the learning process. Once the model is built, it is trained on the training data, with the main goal being to minimize the disparity between the super-resolution images generated by the model and the real high-resolution images present in the dataset. This training phase often requires significant resources in terms of computing power and time. After training, the model is validated on a validation dataset to ensure its generalizability to new images. Metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity

Index (SSIM) are used to assess the quality of the super-resolved images. Example of a current approach: Super-Resolution Convolutional Neural Network (SRCNN) [17], Fast Super-Resolution Convolutional Neural Network (FSRCNN) [20], Very Deep Super-Resolution Network (VDSR) [21], Efficient Sub-Pixel Convolutional Neural Network (ESPNet) [22], Deep Recursive Residual Network (DRRN) [23].

The application of Generative Adversarial Networks (GANs) for image SR is another significant advancement. Because GANs can produce realistic images, they have completely changed the sector and resulted in creative, powerful methods. There are multiple exciting phases to the GAN-based super-resolution method[24].

The generator is a neural network that aims to produce high-resolution images as an output by using low-resolution images as input. A discriminator is also constructed in parallel. Distinguishing between the images produced by the generator and real, high-resolution images is the discriminator's job. The two networks are trained simultaneously and concurrently. During training, the generator constantly tries to improve its ability to confuse the discriminator by generating more and more realistic images. The discriminator, on the other hand, tries to refine its ability to distinguish. This game of opposition between the generator and the discriminator leads to an iterative learning process in which the generator becomes increasingly skilled at generating high-quality, super-resolution images[25].

The training process continues until equilibrium is reached, where the generator produces high-resolution images indistinguishable from real images. At this point, the GAN model is considered successfully trained. Methods such as SRGAN [26], EdgeSRGAN [27], and ESRGAN [28] have been proposed, enabling even better results in terms of visual quality.

Additionally, ESRGAN-DP [29], IRE [30] and EESRGAN [31] have also emerged. Bi-ESRGAN [32] introduces a novel method for document image super-resolution employing deep transfer learning.

Other innovative approaches, including Real-ESRGAN[33], ESRGAN[34], and IESRGAN [35] have made significant strides in the field.

ESRGAN is a powerful tool for improving image quality and resolution. Researchers have developed various variants of ESRGAN for this purpose. These variants differ in their architectures, training methods and ability to solve specific problems, making them applicable to a wide range of scenarios.

Our approach, GA-GESRGAN, is a document imaging-specific variation of ESRGAN. To enhance the clarity and quality of original document images, this novel approach combines layer document extraction with Gabor filters, deep learning with ESRGAN neural networks, and optimization with genetic algorithms. The goal is to convert low-resolution, deteriorated documents images into clear and detailed images.

3. OUR SUPER-RESOLUTION METHOD TO IMPROVE DOCUMENT IMAGE QUALITY

As previously mentioned, the foundation of our strategy is an improvement of the ESRGAN method. The figure below illustrates the steps of the proposed approach:

The proposed SR approach to document images can be described as follows:

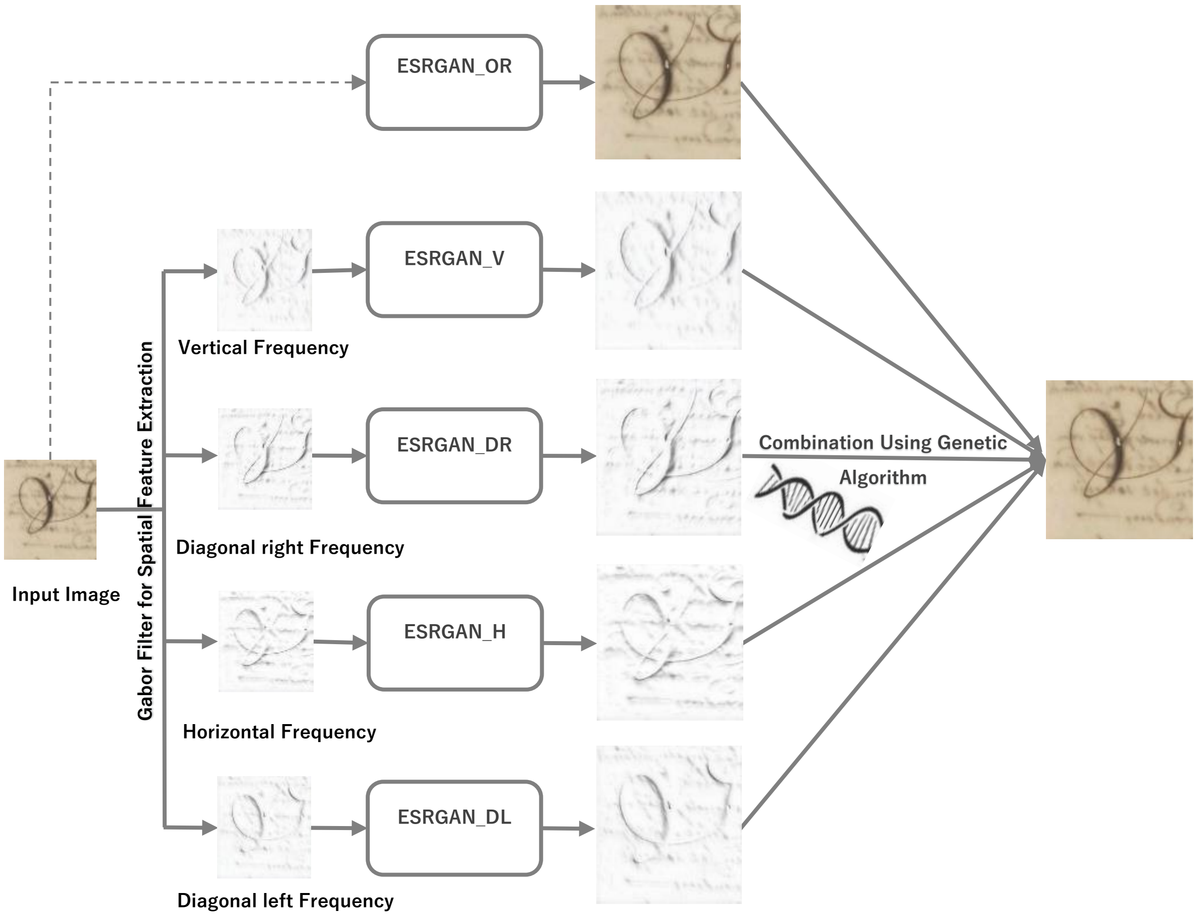

Figure 2 The processes of the proposed SR approach to document images

1.1 GABOR FILTER FOR FEATURE EXTRACTION

The fundamental principle of the Gabor filter is its ability to decompose an image into spatial frequencies and orientations. It detects complex patterns, such as textures, contours, and structures, at different scales and orientations. This capability makes the Gabor filter an ideal choice for tasks such as face detection [36], image segmentation [37], and handwriting recognition [38, 39].

Τηε Γαβορ filter is mathematically defined as follows:

$$G(x,y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{\frac{\tilde{x}^2}{2\sigma_x^2}+\frac{\tilde{y}^2}{2\sigma_y^2}} e^{2\pi j W \tilde{x}} \quad 1$$

$$\begin{cases} \tilde{x} = x\cos\theta + y\sin\theta \\ \tilde{y} = -x\sin\theta + y\cos\theta \end{cases}$$

Where:

(x, y): are the pixels coordinates.

θ: is the orientation of the sinusoid.

W: denotes the radial frequency of the sinusoid.

σ: specifies the scale of the Gabor filter, influencing the effective size of the pixel's neighborhood where weighted summation takes place.

In our investigation, we constrained the number of filters by choosing pertinent directions (θ = 0°, 45°, 90°, and 135°). The selection of these directions stems from their effectiveness in edge recognition, which captures a variety of orientations that are essential for text layouts. It supports all orientations (horizontal, diagonal, and vertical) and is flexible enough to accommodate different document layouts.

Our approach uses the Gabor filter to detect features such as textures and contours and highlight regions of the image, making it a valuable tool for improving the quality of document images. The Figure 3 shows an example of the application of the Gabor filter.

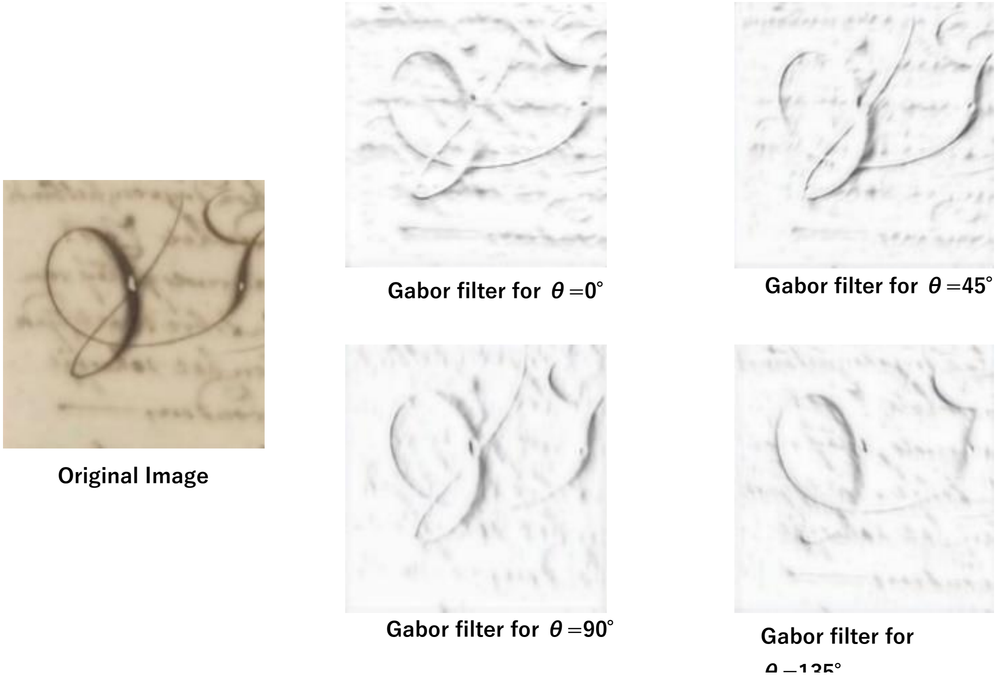

Figure 3 Example of Gabor filter transfer application

This extracted information is then integrated with deep learning techniques, in particular the ESRGAN model. This integration enables us significantly improve image resolution and quality.

1.2 TRANSFER DEEP LEARNING USING ESRGAN

In our methodology, we used the features extracted by the Gabor filter at different spatial frequencies to enrich the training of ESRGAN models. For training the ESRGAN models, we used pairs of training images, where each pair consists of a low-resolution document image and its corresponding high-resolution version. ESRGAN model had been initially trained on nature images from datasets DIV2K, Flickr2K, and ImageNet. We further fine-tuned this model using a specialized database, SR_IVISION_LIMOSE, created for document image super-resolution. The SR_IVISION_LIMOSE dataset was constructed within the Computer Modeling, Optimization, and Electronic Systems Laboratory at the University of Boumerdes. This dataset is designed to support the training and testing of deep learning super-resolution methods specifically for scanned document images.

The SR_IVISION_LIMOSE dataset is organized into two main sections, TEST and TRAIN, with 757 images in each. The dataset was organized into three size categories: high-resolution images at 512x512 pixels, low-resolution at 256x256 pixels, and even smaller low-resolution images at 128x128 pixels. This categorization aids in the evaluation of super-resolution methods.

We used the following parameters for training:

Number of epochs: We chose to form the model over 20 epochs to ensure adequate convergence while avoiding over-learning.

Batch size: We used a batch size of 32 for each learning iteration to ensure good generalization and efficient use of computational resources.

Learning rate: We have initialized the learning rate to 0.001.

Loss function: We used the Mean Squared Error MSE function is used to measure the mean quadratic difference between the values predicted by the model and the true values.

Optimization function: We used Stochastic Gradient Descent SGD is called "stochastic" because it uses a random sample of data at each iteration to estimate the gradient, making training more efficient on large datasets.

By adjusting these parameters, we were able to achieve satisfactory results in terms of

super-resolution quality while maintaining an efficient and robust training of the model.
We trained five ESRGAN models in total:
ESRGAN_OI: An ESRGAN model trained using only the original image pairs as a reference point.
ESRGAN_V: An ESRGAN model trained using the Gabor features corresponding to the vertical frequency components with θ = 90°.
ESRGAN_H: An ESRGAN model trained using Gabor features corresponding to horizontal frequency components with θ = 0°.
ESRGAN_DR: Another ESRGAN model trained using Gabor features corresponding to diagonal right frequency components with θ = 45°.
ESRGAN_DL: Finally, the last ESRGAN model was trained using the Gabor characteristics corresponding to diagonal left frequency components with θ = 135°.
Figure 4 illustrates the results of applying the trained ESRGAN models to a given input image. Each ESRGAN model has as input the image with specific Gabor features (Figure 3), derived from the Gabor filter, and the original image.

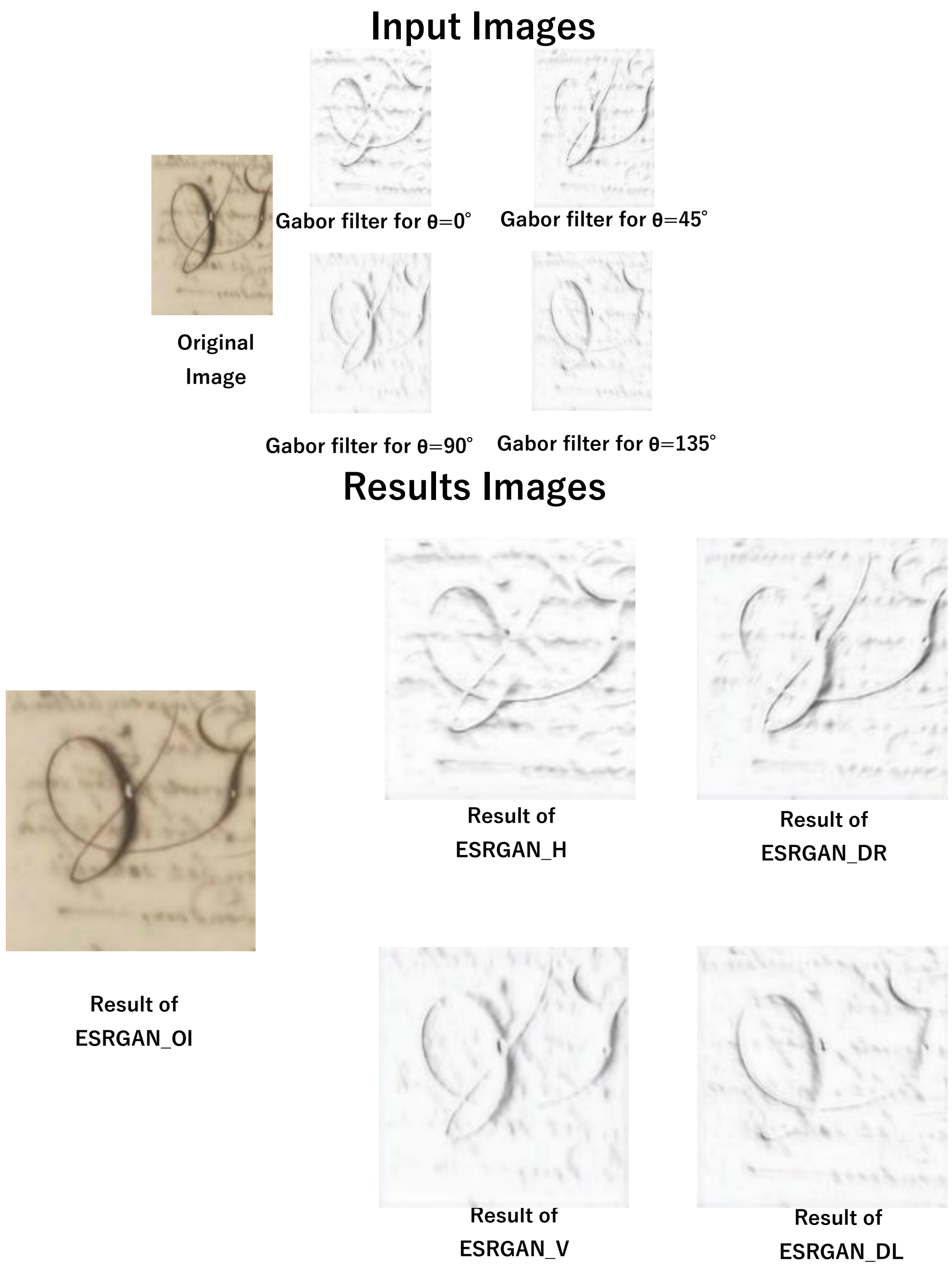

Figure 4 Result of ESRGANs models application

In Figure 4, we can observe the effects and enhancements achieved by each of the five ESRGAN models The combination of the images generated by ESRGAN models through genetic algorithms allows for super-resolution while preserving important details and textures in the input image.

1.3 COMBINATION USING GENETIC ALGORITHMS

The use of a genetic algorithm in our approach is in the particular context of optimizing weights associated with ESRGAN models. This choice is based on several considerations. Firstly, our problem involves the simultaneous optimization of multiple weights that correspond to different versions of the ESRGAN model. With a large search space and parameters to be adjusted consistently, the genetic algorithm offers a systematic and iterative approach to finding optimal solutions.

Genetic algorithms belong to the family of evolutionary algorithms introduced by John Holland (University of Michigan, USA) in the early 1970s and further developed by Golodberg, who used them to solve concrete optimization problems[40]. These algorithms are inspired by biological phenomena and are based on the concept that the best-adapted individuals tend to live long enough to reproduce, while the weakest tend to disappear. This approach has applications in a variety of fields, from parameter optimization to neural network training. [41].

After obtaining images enhanced by ESRGAN models, the method uses a genetic algorithm to combine these results. The object of this algorithm is to determine the optimal pondiration in order to maximize the PSNR. This is done by adjusting the coefficients associated with each of the images generated by the ESRGAN models. Which represents the individual or the solution in a genetic algorithm.

$$Individual = [gene_1, gene_2, \ldots, gene_n]$$

Each element in this vector represents a gene, and the entire vector encodes a potential solution with specific weights assigned to each ESRGAN model.

By adjusting these weights (solution), we aim to optimize the quality of the resulting image, focusing on features such as sharpness and clarity.

A visual depiction of the bioinspired mechanisms that unfold during a single iteration of genetic optimization is represented in the Figure 5:

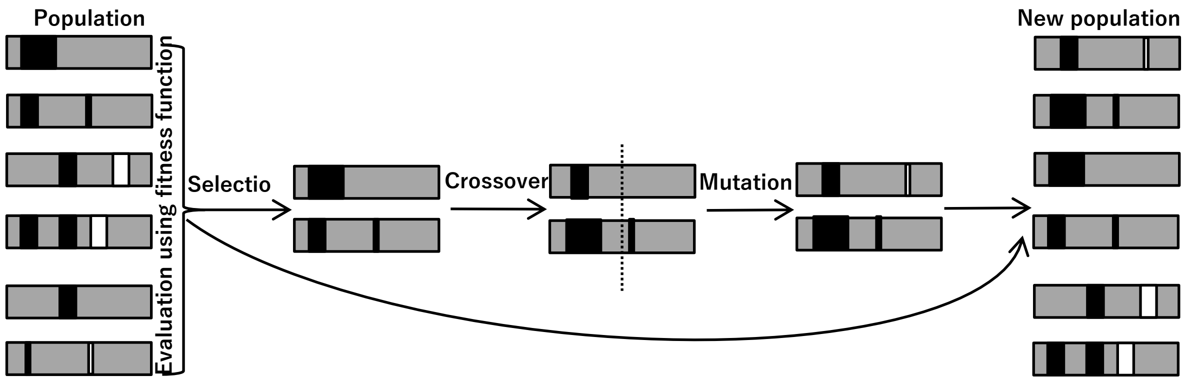

Figure 5 Schematic representation of the genetic algorithm process during one iteration

The algorithm begins by creating an initial population of solutions, which are randomly created to ensure that they are distributed over the entire search space.

Each solution is evaluated in terms of its ability to solve the posed problem. Using a fitness function F, in our case, involves calculating the PSNR. The formula for fitness function F in our method is as follows:

F=PSNR(HR,(Solution[0] * OI + Solution[1] * HI + Solution[2] * DRI+ Solution[3] * VI + Solution[4] * DLI))

Where:

OI, VI, DRI, HI, and DLI are the images results of ESRGAN models, specifically ESRGAN_OI, ESRGAN_V, ESRGAN_DR, ESRGAN_H, and ESRGAN_DL models, respectively.

Solution[i], i in [0, 4]: These are coefficients applied to each of the generated images to combine them into a single image.

The selection phase in a genetic algorithm is crucial in determining which individuals will have the chance to reproduce and thus influence the next generation. Several selection methods exist, each with its own characteristics. Roulette selection assigns selection probabilities according to individual performance. Elitism preserves the best-performing individuals. Tournaments pit randomly selected individuals against each other, while uniform selection gives all individuals the same chance of being chosen. Roulette Wheel Selection the probability of an individual being selected is directly proportional to its fitness score. Individuals with higher fitness have a higher chance of being selected[42].

In our method, we have opted for Roulette Wheel Selection, as it prioritizes the selection of the best-performing individuals, which speeds up the optimization process.

The selected solutions are then combined to create new solutions, often by crossing their genes and introducing random mutations to diversify the population.

The crossover operator in a genetic algorithm is fundamental to the exploration of the search space. It involves combining two parent individuals to create two offspring, called children. These children are not necessarily better than their parents, as the aim is to introduce diversity into the population. The probability of crossover, Pc, determines the frequency with which this operation is applied. Several crossing methods are used, including Mono_point crossing, Multi_point crossing, Uniform crossing, etc.

We opted for our approach of single-point crossover with Pc = 0.8 because of its simplicity of implementation, making our algorithm less complex and easier to manage and understand. It also preserves the genetic structures of the parents over a significant portion of their descendants, which is beneficial for our specific problem.

The mutation operator is used to randomly modify the value of certain genes within an individual. This operation is designed to avoid too rapid a convergence towards a particular individual or a small group of individuals within the population. The probability of mutation, generally noted as Pm, is often lower than that of crossover, Pc. Various types of mutation are commonly used, including: simple mutation, center inversion mutation, reverse mutation, Random mutation, two-gene exchange mutation, three-gene exchange mutation, four-gene exchange mutation...etc.

We have chosen a random mutation with Pm = 0.1 for our approach. Random mutations disrupt individuals by randomly changing some of their genetic characteristics. This prevents the population from converging prematurely to a suboptimal solution by maintaining genetic diversity. This process is repeated over several generations. Each generation is generally more efficient than the last, as the most effective solutions are more likely to be passed on to the next generation. The genetic algorithm runs until a stopping criterion is reached; this condition is either: A maximum number of generations, A minimum fitness value, The population ceases to evolve or no longer evolves sufficiently. In our approach, we utilized a predefined maximum number of iterations as the stopping criterion for the genetic algorithm. The final solution is then returned as an approximation of the best possible solution for the given problem [43].

The next section gives demonstrations of the promising results that have been produced with our approach addressed to the problem of images document super-resolution.

2 RESULTS ANALYSIS AND EVALUATION

Evaluating and analyzing the results is an important part of our method. This section will give a detailed analysis of our model's performance as well as the results of our experiments. We will compare conventional and deep learning-based super-resolution approaches with our GA-GESRGAN method. The benefits of our method will be highlighted through this comparison analysis, particularly with document images (Figs. 6, 7, and Table 1).

The GA-GESRGAN super-resolution model was trained using a comprehensive approach, initially with images from various databases such as DIV2K and ImageNet, then refined specifically with the specialized SR_IVISION_LIMOSE database to capture details and features unique to document images. The use of a loss function MSE and an optimizer SGD with a learning rate of 0.001 allowed to reach a high quality of super-resolution while avoiding overlearning, thus promoting a robust convergence of the model over 20 periods with batches of size 32. These settings ensured efficient generalization and optimal use of resources, providing a reliable solution for super-resolution of document images with preservation of essential details.

To evaluate the quality of the images, we will take into account measures such as optical character recognition (OCR), PSNR, SSIM, Visual Information Fidelity (VIF), and Mean Absolute Error (MAE). High values of PSNR, SSIM and VIF indicate better fidelity and structural similarity between the super-resolved image and the original, while a low value of MAE indicates a smaller pixel-to-pixel difference between the two images. These metrics provide a quantitative assessment of the super-resolved image quality and allow to compare different super-resolution approaches.

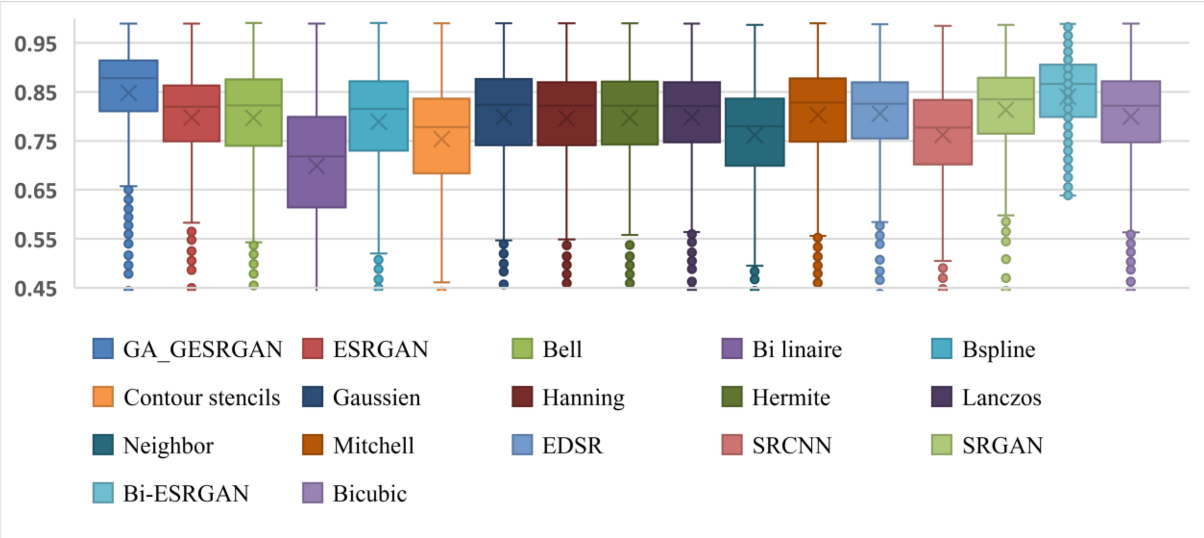

Figure 6 Representative SSIM curves obtained by applying traditional and deep learning-based super-resolution methods on LIMOS database.

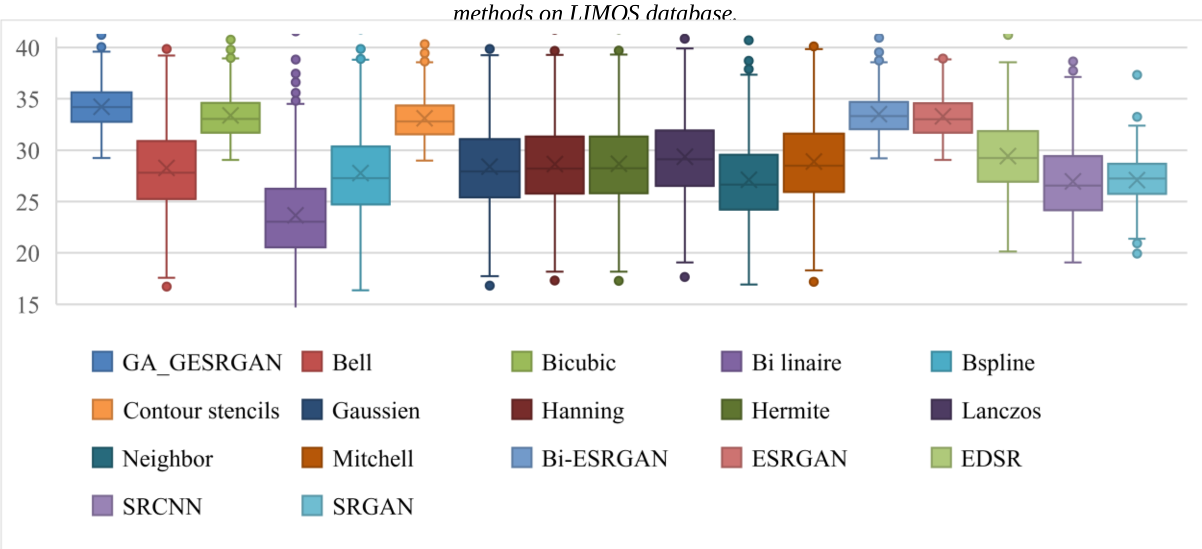

Figure 7 Representative PSNR curves obtained by applying traditional and deep learning-based super-resolution methods on LIMOS database

Table 1 Representative table of PSNR, SSIM, VIF ,and MAE values obtained by applying traditional and deep learning-based super-resolution methods on LIMOS database

Method	PSNR	SSIM	VIF	MAE
GA-GESRGAN	**34.24**	**0.85**	**0.31**	**101.81**
Bell	28.25	0.79	0.22	105.03
Bi linaire	23.64	0.79	0.09	113.04
Bspline	27.75	0.69	0.20	103.20
Contour stencils	33.11	0.78	0.15	100.03
Gaussien	28.38	0.75	0.22	105.45
Hanning	28.64	0.79	0.23	108.79
Hermite	28.66	0.79	0.23	108.66
Lanczos	29.35	0.79	0.24	111.39
Neighbor	27.12	0.79	0.19	112.82
Mitchell	28.88	0.76	0.23	108.28
ESRGAN	33.30	0.80	0.26	116.69
EDSR	29.43	0.80	0.27	116.61
SRCNN	26.94	0.76	0.22	116.69
SRGAN	27.09	0.81	0.26	116.17
Bi-ESRGAN	33.50	0.84	0.29	120.48
Bicubic	33.38	0.79	0.28	153.28

The results of PSNR ,SSIM VIF and MAE values in Figure 6, Figure 7, and Table 1 show that our method, GA-GESRGAN, significantly outperforms several other methods, including SRGAN, SRCNN, EDSR, Bell, Bilinaire, Bspline, contour stencils, Gaussian, Hanning, Her-mite, Lanczos, Neighbor, and Mitchell. This superiority means that our method is more effective at preserving detail and fidelity in relation to the reference image than these competing methods. Among the leading methods, ESRGAN and Bi-ESRGAN also rank highly, although our method performs slightly better. This indicates that our approach manages to reproduce image details more faithfully than these renowned methods.

To evaluate the effectiveness of our method, we also used an OCR metric. OCR is a technique for converting text images into editable text. Using OCR, we measured the ability of our method to improve image quality to the point where text can be extracted more accurately (See Figure 8 and Table 2).

Table 2 Representative table of recognition percentages using OCR

Method	OCR(%)
GA-GESRGAN	**95**
ESRGAN	78
EDSR	75
SRGAN	76
Bi-ESRGAN	89
Bicubic	62

Figure 8 Example of evaluation using OCR

The results in Figure 8 and Table 2 showed that the GA-GESRGAN method was able to produce higher-quality document images, leading to a significant improvement in OCR accuracy. This means that our method can restore the image so that the text can be more easily read and extracted with greater accuracy. We recognize that a purely quantitative assessment based on PSNR and SSIM is not sufficient to fully capture visual quality as perceived by the human eye. Therefore, to obtain a more comprehensive assessment, we performed a visual comparison by presenting several super-resolved images generated by our method alongside those produced by other methods. From Figure9, we see that bicubic interpolation improves resolution little, producing blurry text and poorly defined details. EDSR and SRGAN offer significant improvements with clearer, but sometimes smoothed or artificially textured texts. ESRGAN, meanwhile, achieves a good balance between detail sharpness and realistic textures. While the results clearly demonstrate the superiority of GA-GESRGAN. The fine details and texts are exceptionally clear and realistic, surpassing other methods.

Figure 9Comparison of the efficacy of our method compared to others methods.

We performed a full comparison of our method on the test databases, which included Set5, Set14, BSD100, and Urban100, with the well-established image super-resolution approaches, namely ESRGAN, SRGAN, RCAN, and EDSR, based on the results published in [25], providing a diverse sample of application scenarios. The table below shows the result of this comparison:

Table 3. Representative table of PSNR values obtained by applying ESRGAN, SRGAN, RCAN ,EDSR and our approach, on Set5, Set14, BSD100 and Urban100 databases.

Method	Training dataset	Set5	Set14	BSD100	Urban100	Manga109
SRCNN	291	30.48/0.86	27.50/0.75	26.90/0.71	24.52/0.72	27.58/0.85
EDSR	DIV2K	32.46/0.89	28.80/0.78	27.71/0.74	26.64/0.80	31.02/0.91
RCAN	DIV2K	32.63/0.90	28.87/0.78	27.77/0.74	26.82/0.80	31.22/0.91
ESRGAN	DF2K	32.73/0.90	28.99/0.79	27.85/0.74	27.03/0.81	31.66/0.91
GA-GESRGAN	DF2K+SR_IVISION_LIMOSE	**32.85/0.91**	**30.66/0.84**	**30.72/0.76**	**30.49/0.81**	**31.67/0.92**

.

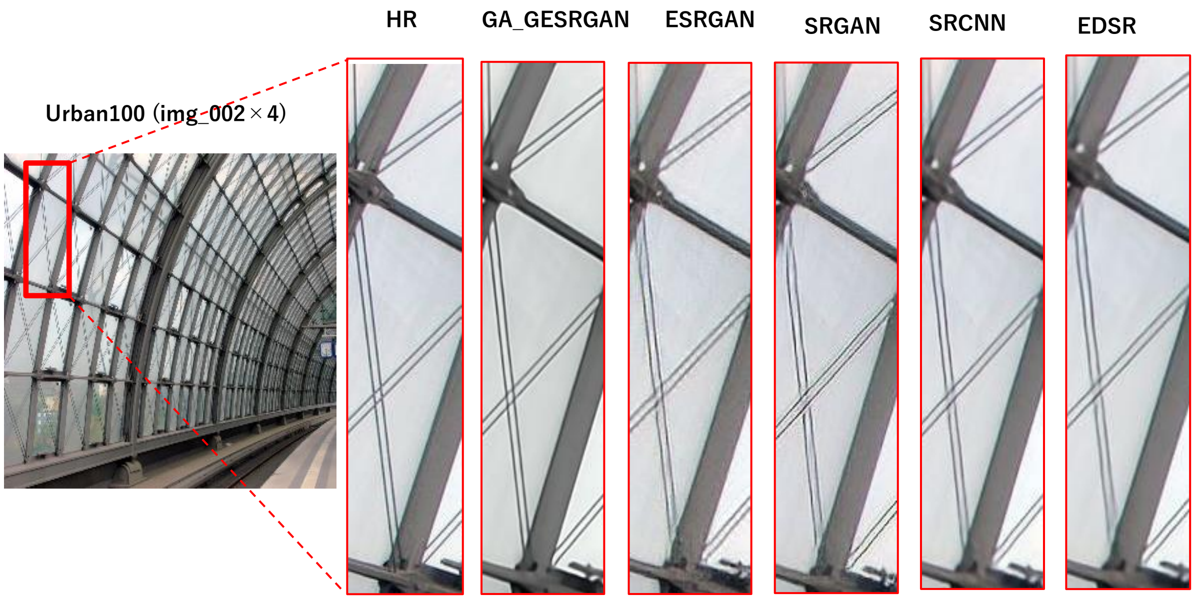

Figure 11 Results of 4× super-resolution on image img_040 from Urban100 Datasets

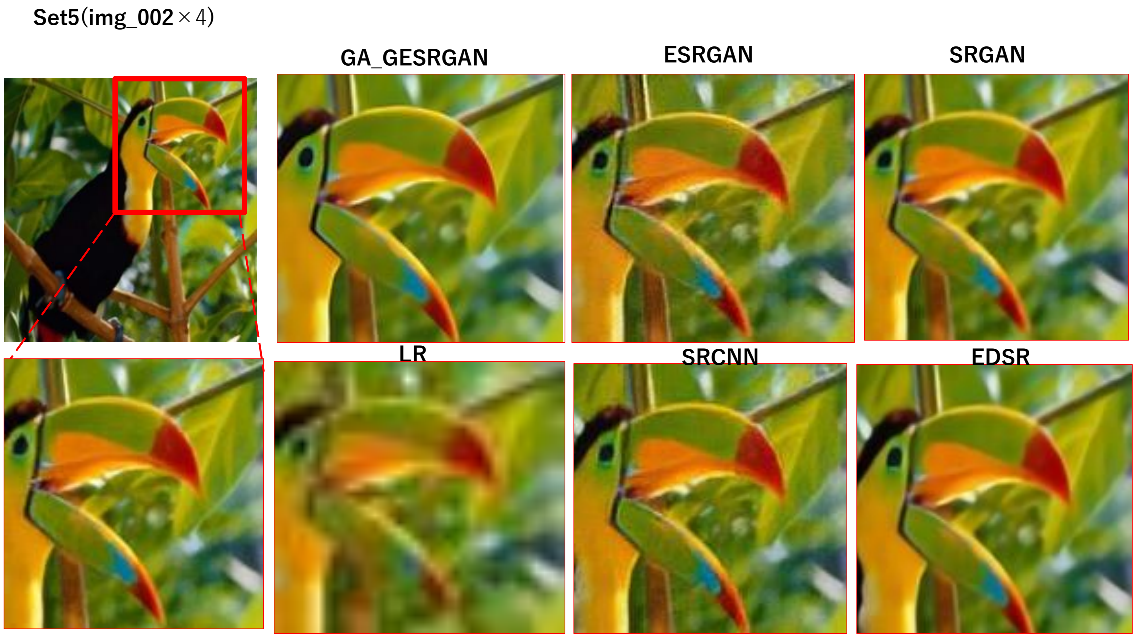

Figure 10 Results of 4× super-resolution on image img_002 from Set5 dataset.

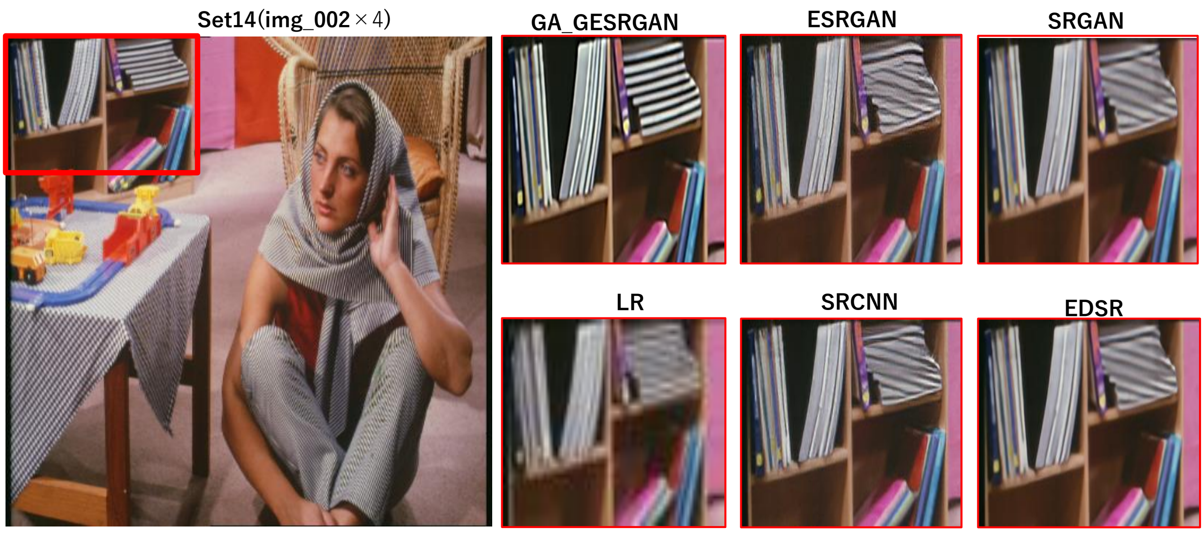

Figure 14 Results of 4× super-resolution on image on img_002 from Set14 dataset

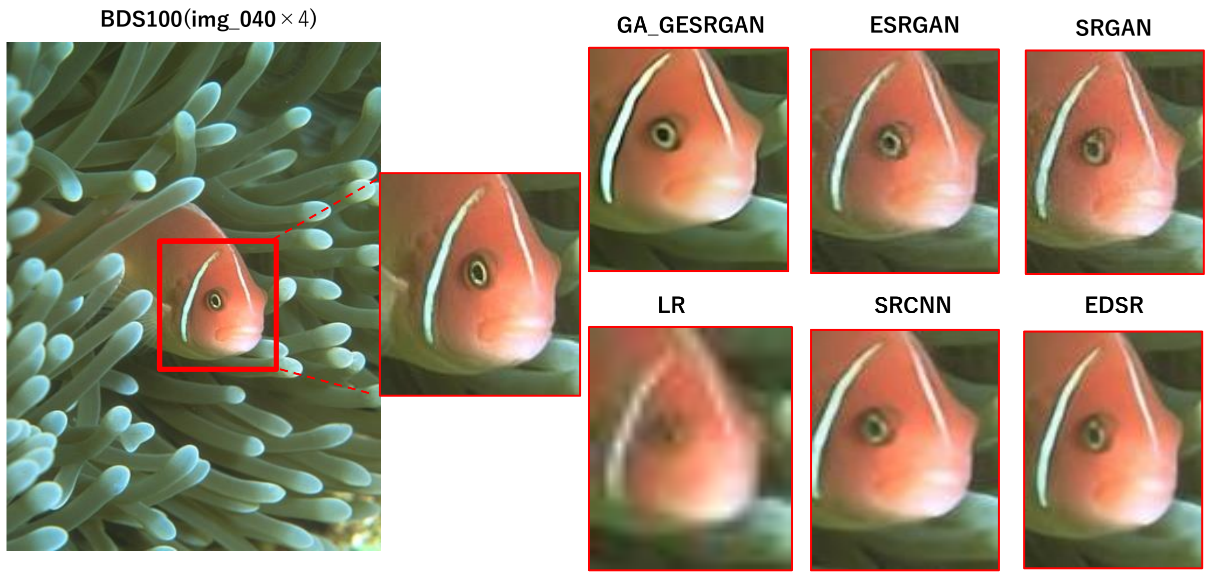

Figure 12 Results of 4× super-resolution on image img_002 from BSD100 dataset

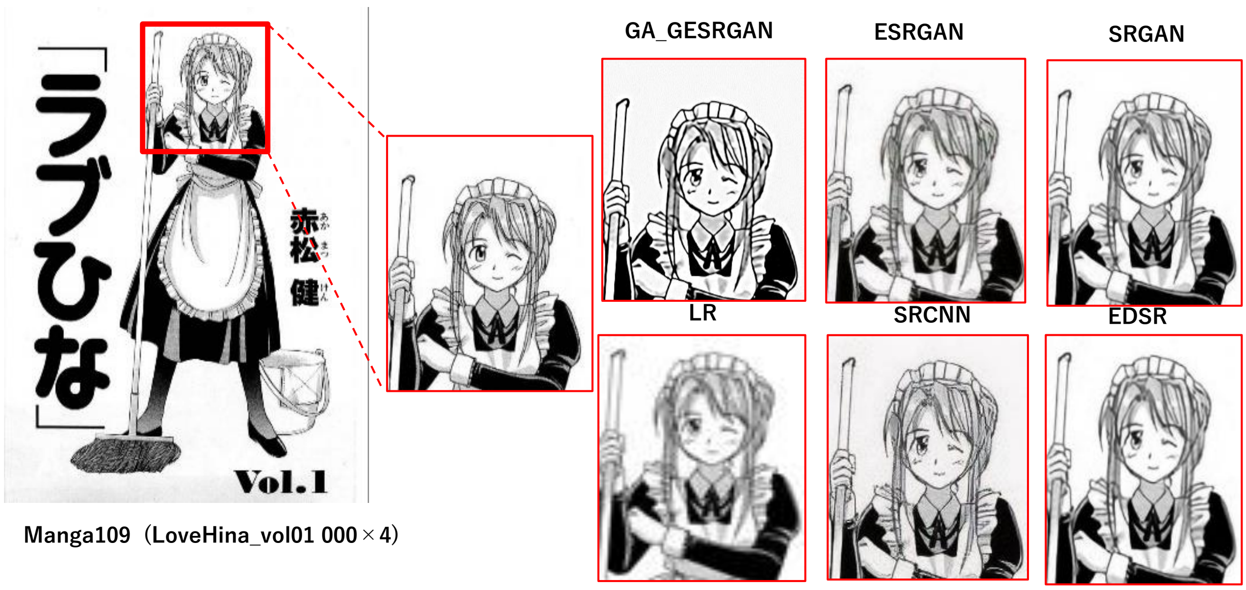

Figure 13 Results of 4× super-resolution on image LoveHina_vol01 000 from Manga109 dataset

We extensively evaluated our super-resolution model performance on a number of popular datasets, including Set5, Set14, BSD100, Urban100, and Manga109. These datasets contain a wide range of situations, from photos taken in the natural world to more specialized genres like manga artwork and urban planning.

The results obtained on the Set5 and Set14 datasets are particularly encouraging. Super-resolved images demonstrate improved visual fidelity, with sharp detail and high structural coherence, as indicated by high PSNR and SSIM values.

Our model constantly shows good performance on BSD100, which has a wide range of complex images. The excellent reproduction of fine textures and contours indicates that the model can generalize well.

An additional challenge is the complex nature of urban structures as represented by Urban100, but this does not seem dependent on how well the model performs. The robustness of our method, which is capable of maintaining high visual quality even in hard instances, is supported by the outcomes.

The results obtained on Manga109 demonstrate that our approach can effectively handle different graphic styles, preserving visual qualities while improving resolution.

This comprehensive evaluation highlights our method's flexibility and adaptability, making it an appealing option for super-resolution image applications in different contexts. Although our method requires a longer run time, its superior performance in terms of quality is solidly confirmed by the results obtained through a wide range of metrics such as PSNR, SSIM, VIF, MAE and OCR, as clearly illustrated in previous tables and graphs. These metrics reveal better visual fidelity, significant noise reduction, increased clarity and accuracy, and improved character recognition, thus justifying the additional execution time required by our method.

Our approach, which takes advantage of Gabor filters for feature extraction, deep learning with the ESRGAN model, and genetic optimization, has been thoroughly evaluated. Results demonstrated a definite advantage in terms of PSNR ,SSIM VIF and MAE. The evaluation relied on OCR was crucial to determining the quality of character recognition. Our results show a real strength by demonstrating how much super-resolved images generated by our method can improve character recognition, which is key to its practicality when text recognition matters. The received visual feedback significantly enriches our comprehension of the effectiveness of the approach, giving crucial details about how users may perceive and interact with the super-resolved results. The application of quantitative and visual methods increases the credibility of our overall evaluation.

3 CONCLUSION

In conclusion, our novel method appears to be superior in terms of super-resolution document images. The use of Gabor filters has improved the quality details recognized by enabling accurate contour and texture recognition. Convincing visual findings were obtained with a considerable improvement in image resolution by the use of the deep learning model, specifically ESRGAN. The genetic algorithm was important in adjusting parameters for the best possible adaptation, which enhanced the overall efficacy of our approach. The combination of visual examination, OCR, and quantitative evaluation using metrics such PSNR and SSIM enhanced the credibility of our approach. This developments pave the way for further improvements, such as additional optimization mechanisms and other deep learning model architectures besides ESRGAN to reveal more optimum combo’s even in super resolution.

ACKNOLEDGMWNT

The authors extend their heartfelt thanks to LIMOSE Laboratory in the Computer Science Department at the University of Boumerdes for providing the opportunity to work on this paper.

REFERENCES

1. Patel, V., Mistree, K. :A review on different image interpolation techniques for image enhancement. International Journal of Emerging Technology and Advanced Engineering, 3(12), 129-133 (2013).
2. Akima, H.: A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points. ACM Transactions on Mathematical Software (TOMS), 4(2), pp. 148-159. (1978).
3. Smith P. R.: Bilinear interpolation of digital images. Ultra microscopy, 6(2), pp. 201-204.(1981).

4. Hall, C. A.: Bicubic interpolation over triangles. Journal of Mathematics and Mechanics,19(1),1-11 (1969).
5. Gordon, W. J. and Riesenfeld, R. F., B-spline curves and surfaces. In : Computer aided geometric design. Academic Press, pp. 95-126 (1974).
6. Gallivan K., Grimme G., and Van Dooren, P.: A rational Lanczos algorithm for model reduction. Numerical Algorithms, 12, pp. 33-63. (1996).
7. Mitchell, D. P. and Netravali A. N. :Reconstruction filters in computer-graphics. ACM Siggraph Computer Graphics, 22(4), pp. 221-228. (1988).
8. Gordon, W. J. and Riesenfeld, R. F.: B-spline curves and surfaces. In : Computer aided geometric design. Academic Press,. pp. 95-126. (1974).
9. Karabassis E. and Spetsakis M. E.: An analysis of image interpolation, differentiation, and reduction using local polynomial fits. Graphical models and image processing, 57(3), pp. 183-196. (1995).
10. Enríquez C., Christian J. et Rodríguez ., Ramón M.: A super-resolution image reconstruction using natural neighbor interpolation. Computación y Sistemas, 19(2), pp. 211-231. (2015).
11. Nguyen D.: Adaptive Windowed Sinc Filter for Image Interpolation. These doctoral. San Jose State University (2023).
12. Hardie R. C. and Barnard K.J.: Fast super-resolution using an adaptive Wiener filter with robustness to local motion. Optics express, 20(19), pp. 21053-21073. (2012).
13. Li X. and Orchard M. T.: New edge-directed interpolation. IEEE transactions on image processing, 10(10), pp. 1521-1527.(2001).
14. Naik, S. and Patel N.: Single Image Super Resolution In Spatial And Wavelet Domain. The International Journal of Multimedia & Its Applications, 5(4), pp. 23. (2013).
15. Kumar N., Rai N. K., and Sethi A.: Learning to predict super resolution wavelet coefficients. In : Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). IEEE, pp. 3468-3471 (2012).
16. Anbarjafari G. and Demirel H.: Image super resolution based on interpolation of wavelet domain high frequency subbands and the spatial domain input image. ETRI journal, 32(3), 390-394 (2010).
17. Dong Ch., Loy Ch. Ch., He K. and al.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2), pp. 295-307. (2015).
18. Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition, (2014).
19. Tai Y., Yang J., and Liu X.: Image super-resolution via deep recursive residual network. In : Proceedings of the IEEE conference on computer vision and pattern recognition, pp 3147-3155. IEEE, (2017).
20. Passarella L. S., Mahajan S., PAL A., and al.: Reconstructing high-resolution ESM data through a novel fast super resolution convolutional neural network (FSRCNN). Geophysical Research Letters, 49(4), pp. e2021GL097571, (2022).
21. Song T., Chowdhury S. R., Kim K., and al.: Super-resolution PET using a very deep convolutional neural network. In: IEEE Nuclear Science Symposium and Medical Imaging Conference Proceedings (NSS/MIC), pp. 1-2. IEEE (2018).
22. Ruan H., Tan Z., Chen L., and al.: Efficient sub-pixel convolutional neural network for terahertz image super-resolution. Optics Letters, 47(12), pp. 3115-3118.(2022)
23. Tai Y., Yang J., and Liu X.: Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 3147-3155, IEEE (2017).
24. Fu, K., Peng, J., Zhang, H., Wang, X., & Jiang, F. Image super-resolution based on generative adversarial networks: a brief review. (2020).
25. Tian, C., Zhang, X., Lin, J. C. W., Zuo, W., Zhang, Y., & Lin, C. W. Generative adversarial networks for image super-resolution: A survey. (2022).
26. Ledig Ch., Theis L., Huszár F.: and al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4681-4690. IEEE (2017).
27. Angarano S., Salvetti F., Martini M., and al.: Generative adversarial super-resolution at the edge with knowledge distillation. Engineering Applications of Artificial Intelligence, vol. 123, pp. 106-107. (2023).
28. Wang X., Yu K., Wu Sh., and al.: Esrgan: Enhanced super-resolution generative adversarial networks. In: Proceedings of the European conference on computer vision (ECCV) workshops, pp. 0-0. Springer (2018)
29. Song J., Yi H., Xu W., and al.: ESRGAN-DP: Enhanced super-resolution generative adversarial network with adaptive dual perceptual loss. Heliyon, 9(4). (2023).

30. Zhu Z., Lei Y., Qin Y., and al.: IRE: Improved Image Super-Resolution Based On Real-ESRGAN. IEEE Access, 2023.
31. Tsai A., Tsou Ch., and Wang J.: EESRGAN: Efficient & Effective Super-Resolution Generative Adversarial Network. IETE Technical Review, pp. 1-12 (2023).
32. Kezzoula Z., Gaceb D., Akli Z., and al.: Bi-ESRGAN: A New Approach of Document Image Super-Resolution Based on Dual Deep Transfer Learning. In : International Conference on Artificial Intelligence: Theories and Applications. Cham : Springer Nature Switzerland, pp. 110-122 (2022).
33. Wang, X., Xie, L., Dong, C., & Shan, Y.: Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In Proceedings of the IEEE/CVF international conference on computer vision, pp. 1905-1914. IEEE (2021).
34. Rakotonirina N. C. and Rasoanaivo A.: ESRGAN+: Further Improving Enhanced Super-Resolution Generative Adversarial Network, ICASSP IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 3637-3641. Barcelona, Spain, (2020).
35. Yue, X., Liu, D., Wang, L., Benediktsson, J. A., Meng, L., & Deng, L: IESRGAN: Enhanced U-Net Structured Generative Adversarial Network for Remote Sensing Image Super-Resolution Reconstruction. Remote Sensing, 15(14), pp:3490 (2023).
36. Rahman, M. T., & Bhuiyan, M. A.: Face recognition using gabor filters. In 11th International Conference on Computer and Information Technology, pp. 510-515. IEEE (2008).
37. Mehrotra, R., Namuduri, K. R., & Ranganathan, N.: Gabor filter-based edge detection. Pattern recognition, 25(12), pp. 1479-1494 (1992).
38. Wang, X., Ding, X., & Liu, C.: Gabor filters-based feature extraction for character recognition. Pattern recognition, 38(3), pp. 369-379 (2005).
39. Gaceb D., Eglin V., Bres S., and al.: Handwriting similarities as features for the characterization of writer's style invariants and image compression. In : Image Analysis and Recognition: Third International Conference, ICIAR 2006, Proceedings, Part II 3, pp. 776-789, Springer, Portugal (2006).
40. Michaud D. : Indexation bio-inspirée pour la recherche d'images par similarité. These doctoral. University OF Poitiers. (2018).
41. Sandeli M. : Traitement d'images par des approches bio-inspirées Application à la segmentation d'images. University of Constantine, (2014).
42. Katoch, S., Chauhan, S. S., & Kumar, V. A review on genetic algorithm: past, present, and future. Multimedia tools and applications, (2021).
43. Bourazza Saïd. : Variantes d'algorithmes génétiques appliquéees aux problèmes d'ordonnancement. These doctoral. University du Havre (2006).