

Beyond a Single Perspective: Neural Fusion of Lévy-Generated Super-Resolution Images for Robust Face Recognition

Marcelo dos Santos*, João C. Neves[†], David Menotti*

*Federal University of Paraná, Curitiba, Brazil

[†]University of Beira Interior, Covilhã, Portugal

*{msantos,menotti}@inf.ufpr.br [†]jcneves@di.ubi.pt

Abstract—Despite significant advancements in facial recognition technology, these systems struggle in real-world surveillance scenarios, especially when dealing with low-resolution images. Super-resolution techniques have emerged as a natural solution to this issue, primarily using generative models such as generative adversarial networks or diffusion models. Despite their potential, these methods face critical limitations, including mode collapse and difficulty in preserving identity-specific features. To address these challenges, we propose Lévy Super-Resolution (LSR), a novel super-resolution approach leveraging Lévy processes as a noise source in a modified diffusion model. This stochastic process enhances diversity and mitigates mode collapse, enabling the generation of diverse super-resolved samples from a single low-resolution input. These samples are fused using an ensemble of neural networks trained with a triplet loss, producing robust face descriptors with reduced bias and improved identity preservation. To the best of our knowledge, LSR is the first method to utilize Lévy processes for super-resolution generation. We validate our approach on real and synthetic low-resolution samples from three standard face recognition datasets, and we demonstrate that our approach can surpass both general and face-specific state-of-the-art super-resolution (SR) methods. Our code is publicly available at <https://github.com/marcelowds/lsr>.

Index Terms—Diffusion Models, Lévy Process, Super-Resolution, Face Recognition

I. INTRODUCTION

Despite the advances in face recognition, state-of-the-art methods still struggle in real-world surveillance scenarios, where face samples are acquired at a distance and on the move. Although these samples are affected by many degradation factors, the low resolution of the data is one of the most critical. To address this problem, the use of SR arises as a natural solution. Even though this idea is not novel, existing approaches still have shortcomings. General SR strategies often fail to preserve identity, resulting in decreased recognition accuracy compared to the use of low-resolution images [1]. On the other hand, most SR methods specifically devised for face generation have focused primarily on enhancing image quality, neglecting the assessment of recognition performance improvements [2], [3]. Existing face super-resolution methods designed for enhancing recognition performance have employed strategies to enforce identity preservation during the training of the generative model (e.g., GANs or Diffusion Models). However, when applied to highly degraded data, these models

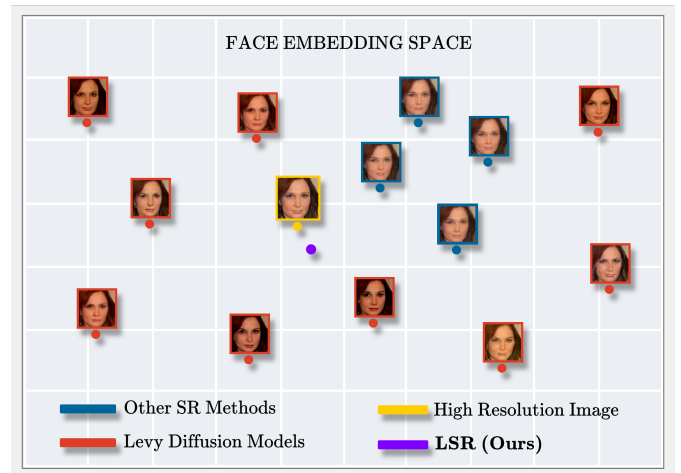


Fig. 1: **Illustration of the distribution of the face embeddings of different super-resolution strategies.** The point • represents the original high-resolution image. Due to the ill-posed nature of the super-resolution problem, traditional SR algorithms produce solutions • around a biased region. The use of a Lévy Diffusion Model can mitigate this by generating more diverse solutions •. These solutions can be combined to reduce bias, producing a closer approximation • to the original face embedding.

are prone to mode collapse, resulting in biased high-resolution approximations that impair recognition performance.

In this paper, we argue that this problem can be addressed by increasing the variance of the generation process while reducing its overall bias through the combination of individual solutions. In particular, we exploit a modified diffusion process capable of generating more diverse super-resolved samples. This flexibility arises from the inherently ill-posed nature of SR problems. As depicted in Fig. 1, diffusion models relying on a Lévy process generate samples whose face embeddings diverge more from the original high-resolution image than other SR methods. However, if these embeddings are properly combined, these samples are less biased towards a specific direction in the face embedding space. Moreover, the generated

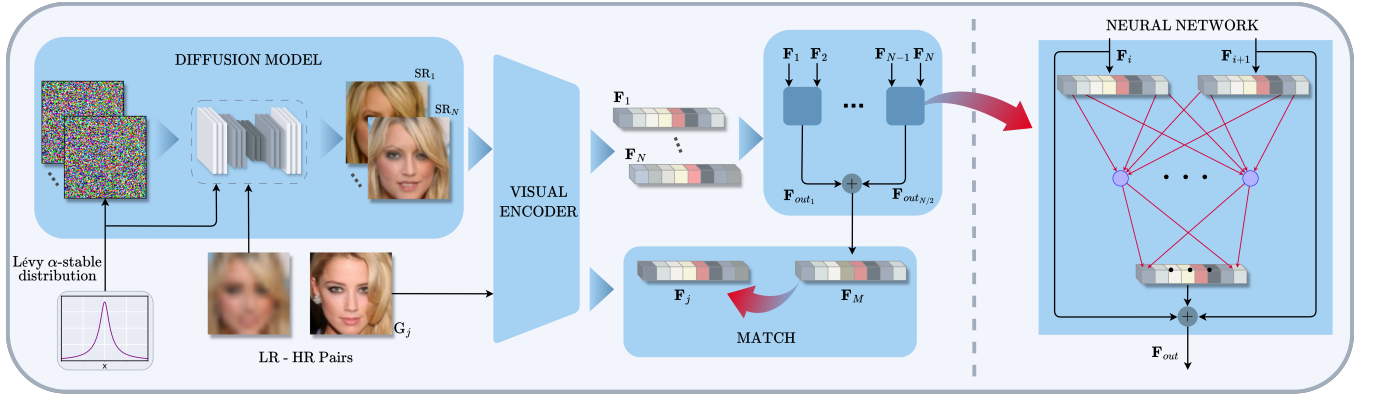


Fig. 2: **Overview of the proposed method.** A low-resolution image (LR) is used to generate a set of super-resolved images ($SR_1, \dots, SR_N$) via a diffusion model based on a Lévy process. Features ($F_1, \dots, F_N$) are extracted from these super-resolution images and combined using an ensemble of neural networks, trained with a triplet loss, to create the merged vector F_M . Subsequently, for facial recognition, this merged vector is compared against the gallery feature set, matching the feature F_j from gallery image G_j , which corresponds to the same identity.

samples by the Lévy process are less susceptible to mode collapse.

Accordingly, we leverage these characteristics by introducing a strategy that combines the embeddings of multiple super-resolved samples to approximate the original high-resolution embedding. To achieve this, an ensemble of neural networks is trained to map N super-resolved face descriptors into a unified descriptor that closely represents the high-resolution image. This approach exploits the hypothesis that aggregating diverse representations reduces noise and enhances fidelity. We evaluated the efficacy of this approach using CelebA [4], QuisCampi [5], and SCface [6] datasets. The proposed approach achieves superior quantitative results against both general and face-specific state-of-the-art SR methods, demonstrating improved performance in face recognition tasks. Additionally, we show that our approach depends only on a small number of super-resolved samples (N) to converge to its maximum performance.

Our contributions are as follows:

- We introduce LSR (Lévy Super-Resolution), a stochastic SR method based on the Lévy-Itô Model (LIM). This approach enhances the diversity of generated faces, allowing the creation of a face descriptor that more accurately approximates the descriptor of the high-resolution (HR) image by utilizing an ensemble of neural networks.
- We extend the metrics used in previous face super-resolution methods by directly assessing recognition accuracy rather than relying solely on the classical identity similarity [7].
- We demonstrate that the proposed method achieves superior recognition performance on synthetic and real-world low-resolution face images.

II. RELATED WORKS

Traditionally, Generative Adversarial Networks (GANs) [8] have dominated image generation tasks. However, training

these models is often unstable, and with the emergence of diffusion models, GANs are being outperformed in various applications [9]. The three pioneering and most significant works in diffusion models are Denoising Diffusion Probabilistic Modeling (DDPM) [10], Score Matching with Langevin Dynamics (SMLD) [11], and Score-Based Generative Modeling through Stochastic Differential Equations (SDEs) [12]. DDPM progressively corrupts the training data to Gaussian Noise and implements a model to reverse each step of the noise corruption, generating new data. SMLD estimates the gradient of the log probability density with respect to data at each noise scale. The samples are generated through a sequence of progressively decreasing noise scales using annealed Langevin dynamics. In [12], the authors introduced a generalization of both SMLD and DDPM to continuous time through the use of SDEs.

The recent advances in diffusion models have been extended to the problem of super-resolution by conditioning the denoising step on the low-resolution sample. However, in spite of the remarkable results obtained by approaches like SR3 [13] and SRDiff [14], these methods have been devised for generic data, and when applied to face images, may distort an individual's identity, thereby compromising the accuracy of facial recognition systems that use SR as a preprocessing step to enhance low-quality images. To address this problem, various approaches have been developed to enforce identity preservation during the image reconstruction process, such as [15] and [16]. However, [15] requires the calculation of gradients during the image reconstruction phase, resulting in increased computational costs. Additionally, both are prone to mode collapse, which can negatively impact facial recognition performance.

A solution to address mode collapse has been recently introduced in [17], where the authors showed that the use of an α -stable Lévy distribution instead of Gaussian noise to define the noise distribution alleviates mode collapse in imbalanced datasets. Based on that work, we hypothesize that

this approach could be exploited to learn a super-resolution diffusion model capable of producing a wider variety of samples while maintaining consistency with the low-resolution (LR) image.

III. PROPOSED METHOD

This section details the proposed LSR model, which relies on Lévy Diffusion Models to increase the diversity of the super-resolved samples, and decrease the bias of the facial feature representations. Considering the importance of the generation process to the overall method, we first briefly introduce α -stable distributions, highlighting some of their properties, followed by an explanation of diffusion models based on these distributions.

A. Stable distribution

A d -dimensional isotropic α -stable distribution, $d \in \mathbb{N}$, is denoted by $S\alpha S^d(\gamma)$. Where γ is a scale parameter and $\alpha \in (0, 2]$ is a characteristic exponent that governs the decay rate of the distribution's tail. An α -stable distribution has a power-law tail, given by $\mathbb{P}(x \geq r) \sim r^{-\alpha}$. In contrast, a Gaussian distribution, commonly used in diffusion models, has an exponentially decaying tail, specifically, $\mathbb{P}(x \geq r) \leq \exp\left(-\frac{r^2}{2\sigma^2}\right)$. Thus, rare events are relatively more frequent for α -stable than Gaussian distributions. The lower the value of α , the heavier the tail of the distribution, and the higher the frequency of large jumps.

Definition III.1 (Lévy Process). *An $\mathbb{R}^d$ -valued stochastic process $\mathbf{L}_t$ is a Lévy process if it satisfies the following properties: (i) $\mathbf{L}_0 = 0$, (ii) $\mathbf{L}_t$ has independent increments, (iii) $\mathbf{L}_t$ has stationary increments and (iv) The sample paths of the process are stochastically continuous. An α -stable Lévy process $\mathbf{L}_t$ satisfies that, for all $s < t$, the increment $\mathbf{L}_t - \mathbf{L}_s = \mathbf{L}_{t-s} \sim S\alpha S^d((t-s)^{1/\alpha})$.*

B. SDEs driven by Lévy processes

Now, we provide a basic understanding of the score-based generative model known as LIM, described in [17], which uses isotropic α -stable Lévy processes. Similar to the Variance Preserving (VP) SDE model defined in [12], Yoon et al. [17] extend the VP formula to an SDE driven by an isotropic α -stable Lévy process:

$$d\mathbf{X}_t = -\frac{\beta(t)}{\alpha} \mathbf{X}_t dt + \beta(t)^{1/\alpha} d\mathbf{L}_t^\alpha. \quad (1)$$

Since we are adding multiple α -stable distributions with the same parameter α , the solution $\mathbf{x}(t)$ takes the form $a(t)\mathbf{x}(0) + \gamma(t)\epsilon$, where $\epsilon \sim S\alpha S^d(1)$, $a(t) = \exp(-\int_0^t \frac{\beta(s)}{\alpha} ds)$, and $\gamma(t) = (1 - a(t))^{1/\alpha}$. All the noise of the reverse diffusion can be simulated by only one term L_t^α since the other terms are comparatively negligible. In this case, the reverse-time SDE is given by

$$d\mathbf{X}_t = \left(-\frac{\beta(t)}{\alpha} \mathbf{X}_t - \alpha \beta(t) S_t^\alpha(\mathbf{x}) \right) dt + \beta^{1/\alpha}(t) d\mathbf{L}_t^\alpha, \quad (2)$$

where S_t^α is the fractional score function defined in [17] as

$$S_t^\alpha(\mathbf{x}) = \frac{\Delta^{\frac{\alpha-2}{2}} \nabla p_t(\mathbf{x})}{p_t(\mathbf{x})}, \quad (3)$$

where p_t is the probability density of $\mathbf{x}$ and $\Delta^{\gamma/2}$ is a fractional laplacian given by

$$\Delta^{\gamma/2} f(\mathbf{x}) = \mathcal{F}^{-1} \{ \|\mathbf{u}\|^\gamma \mathcal{F}\{f\}(\mathbf{u}) \}. \quad (4)$$

$\mathcal{F}$ and $\mathcal{F}^{-1}$ are the Fourier transform and its inverse, respectively. Using Euler-Maruyama, (2) becomes [17]

$$\mathbf{x}_t = \rho \mathbf{x}_s + \alpha^2 (\rho - 1) S_s^{(\alpha)}(\mathbf{x}_s) + (\rho^\alpha - 1)^{1/\alpha} \epsilon, \quad (5)$$

where ρ is the ratio $a(t)/a(s)$, $\epsilon \sim S\alpha S^d(1)$ for $s > t$, and $\Delta t = s - t$ is a small timestep.

To solve (5), we need an approximation for the fractional score function $S_t^{(\alpha)}(\mathbf{x}_t)$. To achieve this, we train a neural network $\mathbf{s}(\mathbf{x}, \mathbf{y}; \theta)$ with parameters θ , which takes as input a noisy image $\mathbf{x}_t$ and a LR image $\mathbf{y}$. These two images are concatenated along the channel dimension. To adjust the network's parameters, the following loss function must be minimized:

$$L_1(\theta) = \mathbb{E}_{t, \mathbf{x}_t \sim p_t(\mathbf{x}_t)} \left\| \mathbf{s}_t(\mathbf{x}_t, \mathbf{y}; \theta) - S_t^{(\alpha)}(\mathbf{x}_t) \right\|_2^2. \quad (6)$$

However, estimating the fractional score function $S_t^{(\alpha)}(\mathbf{x}_t)$ is challenging. Yoon et al. [17] demonstrated that minimizing L_1 is equivalent to minimizing the conditional score matching loss function, defined as:

$$L_2(\theta) = \mathbb{E}_{t, (\mathbf{x}_0, \mathbf{x}_t) \sim p_t(\mathbf{x}_0, \mathbf{x}_t)} \left\| \mathbf{s}_t(\mathbf{x}_t, \mathbf{y}; \theta) - S_t^{(\alpha)}(\mathbf{x}_t | \mathbf{x}_0) \right\|_2^2. \quad (7)$$

In order to analytically derive $S_t^{(\alpha)}(\mathbf{x}_t | \mathbf{x}_0)$ and thus evaluate L_2 , we first need to express the underlying distribution involved in the transition. Let $q_\alpha(\mathbf{x})$ be the density of $S\alpha S^d(1)$, then we have the following transition kernel from $\mathbf{x}_0$ to $\mathbf{x}_t$ [17]:

$$p_t(\mathbf{x}_t | \mathbf{x}_0) = \frac{1}{\gamma^d(t)} q_\alpha \left(\frac{\mathbf{x}_t - a(t)\mathbf{x}_0}{\gamma(t)} \right). \quad (8)$$

With $p_t(\mathbf{x}_t | \mathbf{x}_0)$, the loss L_2 can be efficiently determined by analytically deriving $S_t^{(\alpha)}(\mathbf{x}_t | \mathbf{x}_0)$, which is done through the following theorem:

Theorem 1 (Yoon et al.). *Let $1 < \alpha < 2$. If the transition density function $p(\mathbf{x}_t | \mathbf{x}_0)$ follows 8, then the fractional score function $S_t^{(\alpha)}(\mathbf{x}_t | \mathbf{x}_0)$ in 7 is given by*

$$S_t^{(\alpha)}(\mathbf{x}_t | \mathbf{x}_0) = -\frac{1}{\alpha \gamma^{\alpha-1}(t)} \left(\frac{\mathbf{x}_t - a(t)\mathbf{x}_0}{\gamma(t)} \right). \quad (9)$$

As in [17], we choose the beta schedule as

$$\beta(t) = -\alpha \frac{d}{dt} \log(\cos(\pi t/2)) \approx \frac{\alpha}{1+r} \frac{\pi}{2} \tan\left(\frac{\pi}{2} \frac{t+r}{1+r}\right), \quad (10)$$

for $0 < r \ll 1$. In this case, $a(t) = \cos(\frac{t+r}{1+r} \frac{\pi}{2}) / \cos(\frac{r}{1+r} \frac{\pi}{2})$. Given Theorem 1 and the loss function defined in (7), the final loss used for training the SR model is

$$\mathbb{E}_{t \sim U(0,1)} \mathbb{E}_{\mathbf{x}_0 \sim p_{data}} \mathbb{E}_{\epsilon \sim S\alpha S^d(1)} \left\| \gamma^{\alpha-1}(t) \cdot \mathbf{s}_t(\mathbf{x}_t, \mathbf{y}; \theta) + \frac{\epsilon}{\alpha} \right\|_2^2. \quad (11)$$

By optimizing the above loss function, $\mathbf{s}_t(\mathbf{x}_t, \mathbf{y}; \theta)$ approximates $S_t^{(\alpha)}(\mathbf{x}_t)$, enabling us to solve (5). We begin with an

image $\mathbf{x}_T$ composed purely of noise, and as a result, we obtain the SR image $\mathbf{x}_0$, whose reconstruction is guided by the LR image $\mathbf{y}$. Due to the stochasticity of diffusion models and the ill-posed nature of the problem, generating an SR image for a given LR image results in slightly different solutions each time. This is illustrated in Fig. 3. We take advantage of this diversity to produce a more discriminative face representation.

C. Combining SR images

Images acquired in real-world surveillance scenarios are primarily composed of low-frequency information. In the high-frequency components, the signal-to-noise ratio can be relatively small. To prevent noise from being disproportionately high in the high-frequency components, a Gaussian low-pass filter is applied to attenuate this noise.

For a given low-resolution (LR) image $\mathbf{y}$, we generate $N = 30$ SR images, and extract a compact face descriptor using a pre-trained visual encoder. Images are combined at feature-level by using an ensemble of neural networks to map the compact descriptors into an approximation of the high-resolution descriptor.

Each neural network takes as input the concatenation of two features, $\mathbf{F}_i$ and $\mathbf{F}_{i+1}$, and outputs a single merged feature. The architecture consists of two layers, ensuring an effective fusion of the input features into a unified representation. Given that the arithmetic mean of the two input features serves as a natural baseline for merging, the network is trained to refine this operation. Specifically, the output $\mathbf{F}_{\text{out}}$ is

$$\mathbf{F}_{\text{out}} = \frac{\mathbf{F}_i + \mathbf{F}_{i+1}}{2} + \delta(\text{concat}(\mathbf{F}_i, \mathbf{F}_{i+1})). \quad (12)$$

The final merged feature vector is the arithmetic mean of all $N/2$ merged features, $\mathbf{F}_{\text{out}}$. For training, we selected 2,520 identities from the CelebA and SCface datasets, with five SR images for each identity. We employ a triplet loss to increase the probability of positive matches while decreasing that of negative matches. In this setup, the anchor sample is $\mathbf{F}_{\text{out}}$, the positive sample is the feature from the ground truth HR image, and the negative sample is a feature from an SR image originating from a different identity. For SCface only, due to the small number of identities, δ -training and LSR testing share identities, but do not share images.

IV. EXPERIMENTS AND RESULTS

A. Experiments

This study utilized four datasets: FFHQ [18] for training, and CelebA, Quis-Campi, and SCface for testing. In our training strategy, LR images were generated by downsampling HR images from the FFHQ dataset by a factor of $8\times$ followed by upsampling using bicubic interpolation to 128×128 . We also employed an alternative training strategy to evaluate our model in real LR data, which is more complex and composed of a combination of degradations. This strategy involved creating LR images by applying random Gaussian blur, adding Gaussian noise, and random adjustments to brightness, contrast, and saturation. Finally, we randomly downsampled the images by

factors ranging from $2\times$ to $4\times$ followed by upsampling back to 128×128 pixels using bicubic interpolation. The resulting LR-HR pairs were used to train the model. The proposed approaches were evaluated on both synthetic and real low-resolution data. Synthetic data were obtained by downsampling images from CelebA and Quis-Campi. For CelebA, we selected 500 identities and chose two images per identity: one gallery image and one probe image. For Quis-Campi, we adopted the ICB-RW challenge setting with 90 identities and five images per identity. SCface was used to evaluate the alternative proposed approach on real data, where low-resolution samples were acquired by surveillance cameras. We selected only the images taken at the farthest distance (4.20 meters) due to their increased difficulty. Testing involved 130 identities and five camera types (cam1 to cam5). Quis-Campi and SCface provide high-resolution gallery images, which we used as the gallery in our experiments.

To extract features from the images, we adopted a state-of-the-art face recognition model AdaFace [19], which generates a 512-dimensional descriptor of the face. The cosine similarity metric was utilized to compare these image descriptors. We benchmarked our proposed algorithm against state-of-the-art super-resolution algorithms, including SR3 [13], GFP-GAN [2], SPARNet [7], SwinIR [20], and FASR [16].

B. Results

Table I presents the results of our proposed method on the CelebA and Quis-Campi datasets. As shown, LSR outperforms all other methods in recognition metrics (Rank-1, Rank-5, and Rank-10), demonstrating an improvement of approximately 4.4 percentage points in Rank-1 over FASR on the CelebA dataset and 3.6 percentage points on the Quis-Campi dataset. Additionally, our method surpasses FASR in the AUC metric for the CelebA dataset. It is important to note that FASR utilizes a set of LR images to generate the SR image, whereas our method requires only a single LR image. Furthermore, merging the data in the embedding space proves to be an effective strategy for recovering information from LR images. Qualitative results from the CelebA dataset are illustrated in Fig. 3.

Table II presents the results for the SCface dataset using a training strategy tailored for data with real-world degradations. In some cases, relying solely on the LR image yields better outcomes than other SR methods. Although these SR approaches produce high-resolution images, the results are often unrealistic and can distort identity, leading to information loss [1]. In contrast, our training strategy is conservative to avoid distortions in identity. It acts more as a denoiser and image enhancer rather than solely focusing on increasing resolution. As a result, our LSR method achieves superior performance compared to both the LR baseline and other SR techniques.

C. Ablation Study

Table III presents the AUC and Rank-1 results obtained on the CelebA dataset, computed using gallery images. Additionally, we report the average identity similarity IS as in

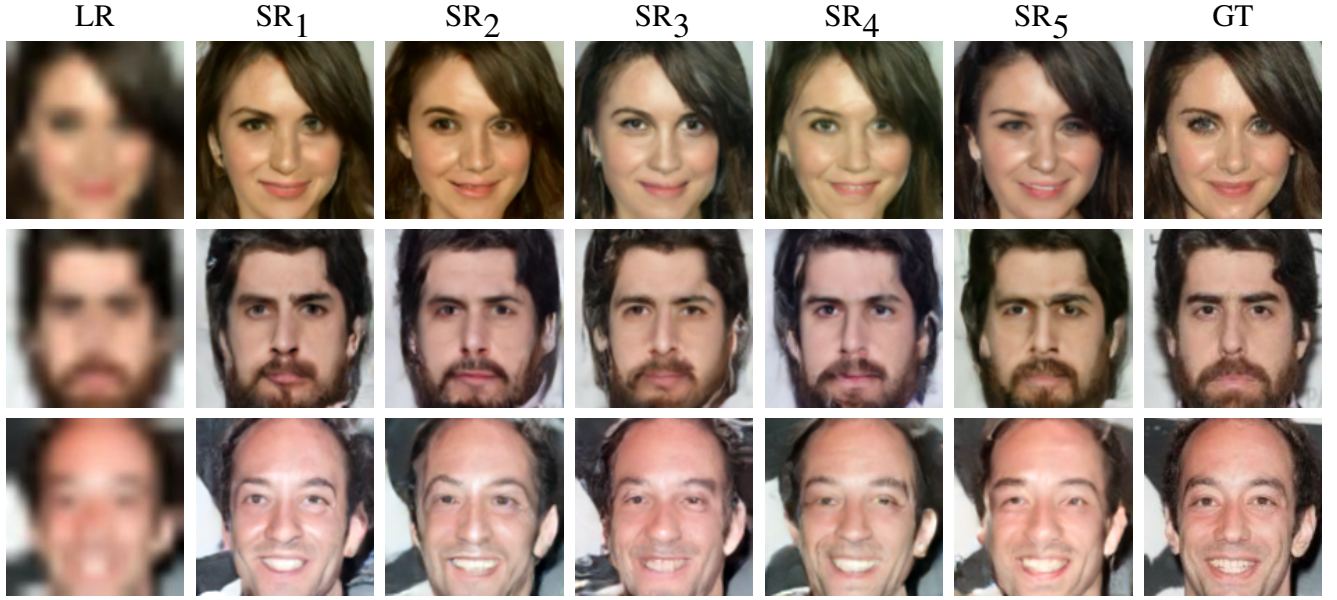


Fig. 3: **Visual results of the Lévy Diffusion Model.** The input low-resolution (LR) image is displayed alongside five generated super-resolved (SR) images ($SR_1, \dots, SR_5$) using an α -stable Lévy distribution, as well as the corresponding ground truth high-resolution (HR) image. The variations in the generated faces result in distinct face embeddings, which, when combined, provide a more discriminative embedding. Additional examples are provided in the appendix section.

TABLE I: Results of 1:1 Verification and 1:N Identification obtained using the AdaFace model (IR-50) for LSR and other SR methods on the CelebA and Quis-Campi datasets.

CelebA				
SR Method	AUC	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
LR	0.915	28.40	50.80	62.20
GFPGAN	0.874	20.60	39.00	46.40
SPARNet	0.875	24.80	41.00	49.20
SR3	0.944	48.00	67.20	72.60
FASR	0.945	56.60	71.20	77.20
LSR (Ours)	0.949	61.00	74.00	79.00

Quis-Campi				
SR Method	AUC	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
LR	0.847	32.67	58.67	69.56
GFPGAN	0.782	20.89	42.00	52.89
SPARNet	0.893	37.11	64.00	74.00
SR3	0.930	50.00	74.89	82.89
FASR	0.941	56.67	78.44	85.56
LSR (Ours)	0.941	60.22	82.44	86.89

TABLE II: Results of 1:1 Verification and 1:N Identification obtained using the AdaFace model (IR-101) for LSR and other SR methods on the SCface dataset.

SCface				
SR Method	AUC	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
LR	0.995	87.54	96.46	98.77
SR3	0.917	40.15	62.46	71.85
SPARNet	0.951	52.00	73.69	83.08
GFPGAN	0.926	42.77	66.77	78.00
SwinIR	0.952	52.62	75.54	82.92
LSR (Ours)	0.997	92.00	98.15	99.54

[7], [21], and the average bias B between SR images and the ground truth HR probe images. The bias is computed by $B = \sum_{i=1}^{|D|} \|\mathbf{F}^{(i)} - \hat{\mathbf{F}}^{(i)}\|_1 / |D|$, where $\mathbf{F}^{(i)}$ is the feature vector derived from the SR probe image(s) associated with identity i , $\hat{\mathbf{F}}^{(i)}$ is the feature vector from the corresponding HR image, and $|D|$ is the size of the test dataset. To facilitate comparison, we normalized the bias by the highest value of this metric in Table III. Furthermore, we evaluate different values of α associated with a Lévy distribution to investigate how its heavy-tailed nature affects the results.

To analyze the impact of the ensemble neural networks on embedding fusion, we compared our method against a scenario where embeddings are combined solely by the arithmetic mean

TABLE III: Impact of the embedding aggregation strategy (LSR vs LSR[†])¹ and tail decay rate (α).

SR Method	α	AUC $\uparrow$	Rank-1 (%) $\uparrow$	IS $\uparrow$	B $\downarrow$
SR3	2.0	0.944	48.00	0.481	1.000
FASR	2.0	0.945	56.60	0.513	0.967
LSR [†]	1.5	0.943	53.00	0.497	0.983
LSR [†]	1.6	0.943	54.20	0.505	0.976
LSR [†]	1.7	0.941	56.00	0.508	0.972
LSR [†]	1.8	0.947	57.60	0.520	0.960
LSR [†]	1.9	0.945	57.60	0.519	0.961
LSR	1.5	0.946	55.00	0.510	0.970
LSR	1.6	0.946	57.40	0.518	0.963
LSR	1.7	0.945	58.20	0.522	0.959
LSR	1.8	0.949	61.00	0.532	0.948
LSR	1.9	0.948	58.80	0.530	0.950

¹ LSR[†] combines embeddings using only the arithmetic mean.

(we denote this baseline as $\text{LSR}^\dagger$). We also compare our approach with other two baselines that use a normal distribution as the noise source: SR3 and FASR. The latter uses features from multiple LR images to reconstruct a single SR image. The results demonstrate that the proposed strategy performs an effective and accurate fusion of face embeddings, which translates into superior results. Moreover, for various values of α , both methods, LSR and $\text{LSR}^\dagger$, consistently achieve superior results compared to the baselines SR3 and FASR. In particular, we observe that for $\alpha = 1.8$, the ensemble of solutions, when combined through the network δ , yields a result with higher AUC, accuracy, identity similarity, and lower bias when compared to the baselines SR3 and FASR. The superior performance of LSR, at $\alpha = 1.8$, indicates that we achieve a balance between diversity and reliability in identity.

D. Accuracy $\times$ Number of Images

We analyze how the number of SR images impacts the discriminability of the face embeddings produced by our LSR method. For this, we randomly sample N images from a set of 50 SR images and repeat this process 10 times, calculating the average accuracy. Fig. 4 illustrates the recognition accuracy as a function of N using two approaches for combining the embeddings of the super-resolved images. The results indicate that the accuracy increases rapidly and asymptotically approaches a maximum value, suggesting that a relatively small number of SR samples is sufficient to achieve quasi-optimal performance.

V. CONCLUSIONS

In this work, we introduced LSR, a SR algorithm based on diffusion models utilizing α -stable Lévy processes. For a given LR input, we generate multiple super-resolved images and combine their embeddings through an ensemble of neural networks, generating a robust face descriptor with reduced bias and higher identity similarity. Additionally, we demonstrate that the proposed approach attains quasi-optimal performance with a small number of super-resolved samples. These findings emphasize the advantages of employing heavy-tailed distributions to improve diversity and quality in super-resolution, achieving state-of-the-art performance on the CelebA, Quis-Campi, and SCface datasets.

ACKNOWLEDGMENTS

This study was financed by: the *Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES)*, Finance Code 001 and the *Programa de Excelência Acadêmica (PROEX)*; the *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)* (# 315409/2023-1).

REFERENCES

- [1] Muskan Dosi, Udaybhan Rathore, Chiranjeev Chiranjeev, Akshay Agarwal, Richa Singh, and Mayank Vatsa, "Is face super resolution truly pushing the boundaries of face recognition?," in *IJCB*, 2024.
- [2] Xintao Wang, Yu Li, Honglun Zhang, and Ying Shan, "Towards real-world blind face restoration with generative facial prior," in *CVPR*, 2021, pp. 9164–9174.

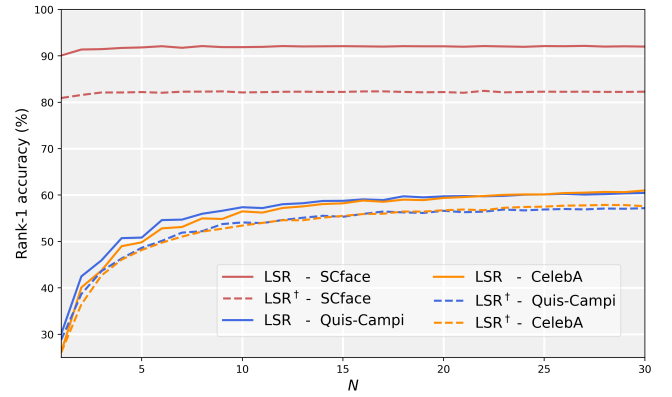


Fig. 4: **Impact of the number of SR images on the LSR rank-1 accuracy.** Accuracy rapidly increases with the number of SR images (N) and asymptotically approaches a maximum.

- [3] Junjun Jiang, Chenyang Wang, Xianming Liu, and Jiayi Ma, "Deep learning-based face super-resolution: A survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–36, 2021.
- [4] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Deep learning face attributes in the wild," in *ICCV*, 2015.
- [5] Joao Neves, Juan Moreno, and Hugo Proença, "QUIS-CAMPI: an annotated multi-biometrics data feed from surveillance scenarios," *IET Biometrics*, vol. 7, no. 4, pp. 371–379, 2018.
- [6] Mislav Grgic, Kresimir Delac, and Sonja Grgic, "SCface – surveillance cameras face database," *Multimedia tools and applications*, 2011.
- [7] Chaofeng Chen, Dihong Gong, Hao Wang, Zhifeng Li, and Kwan-Yee K Wong, "Learning spatial attention for face super-resolution," *IEEE Transactions on Image Processing*, vol. 30, pp. 1219–1231, 2020.
- [8] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *NeurIPS*, 2014, pp. 2672–2680.
- [9] Prafulla Dhariwal and Alexander Nichol, "Diffusion models beat GANs on image synthesis," in *International Conference on Neural Information Processing Systems (NeurIPS)*, 2021, pp. 8780–8794.
- [10] Jonathan Ho, Ajay Jain, and Pieter Abbeel, "Denoising diffusion probabilistic models," in *International Conference on Neural Information Processing Systems (NeurIPS)*, 2020, vol. 33, pp. 6840–6851.
- [11] Yang Song and Stefano Ermon, "Generative modeling by estimating gradients of the data distribution," in *International Conference on Neural Information Processing Systems (NeurIPS)*, 2019, pp. 1–13.
- [12] Yang Song et al., "Score-based generative modeling through stochastic differential equations," in *ICLR*, May 2021, pp. 1–36.
- [13] Chitwan Saharia et al., "Image super-resolution via iterative refinement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 4, pp. 4713–4726, 2023.
- [14] Haoying Li, Yifan Yang, Meng Chang, Shiqi Chen, Huajun Feng, Zhihai Xu, Qi Li, and Yueting Chen, "SRDiff: Single image super-resolution with diffusion probabilistic models," *Neurocomputing*, 2022.
- [15] Maitreya Suin et al., "Diffuse and restore: A region-adaptive diffusion model for identity-preserving blind face restoration," in *WACV*, 2024.
- [16] M. dos Santos, R. Laroca, R. O. Ribeiro, J. Neves, and D. Menotti, "Multi-feature aggregation in diffusion models for enhanced face super-resolution," in *SIBGRAPI*, 2024.
- [17] Eun Bi Yoon, Keehun Park, Sungwoong Kim, and Sungbin Lim, "Score-based generative models with lévy processes," in *NeurIPS*, 2023.
- [18] Tero Karras, Samuli Laine, and Timo Aila, "A style-based generator architecture for generative adversarial networks," in *CVPR*, 2019.
- [19] Minchul Kim, Anil K Jain, and Xiaoming Liu, "AdaFace: Quality adaptive margin for face recognition," in *CVPR*, 2022.
- [20] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte, "Swinir: Image restoration using swin transformer," in *ICCVW*, 2021.
- [21] Kaipeng Zhang, Zhanpeng Zhang, Chia-Wen Cheng, Winston H Hsu, Yu Qiao, Wei Liu, and Tong Zhang, "Super-identity convolutional neural network for face hallucination," in *ECCV*, 2018, pp. 183–198.

APPENDIX

In this paper, we introduced Lévy Super-Resolution (LSR), an algorithm that utilizes Lévy α -stable distributions as a noise source in a diffusion model. This stochastic approach provides image diversity and prevents mode collapse, enabling the generation of multiple diverse super-resolved samples from a single low-resolution input. The features from these images are combined through an ensemble of neural networks, providing a robust image descriptor for face recognition. Figs. 5 and 6 provide additional examples of super-resolved images generated using LSR. The displayed images include low-resolution images, super-resolved images $SR_1, \dots, SR_5$, and the ground truth high-resolution images.

In addition to the results presented in the paper for the SCface dataset, Table IV provides the results of 1:1 Verification and 1:N Identification (Rank-1, Rank-5, and Rank-10) obtained using the AdaFace model (IR-50). Specifically, our LSR method demonstrates an improvement of approximately 12.15 percentage points in Rank-1 compared to LR images, highlighting its effectiveness in real-world scenarios.

TABLE IV: Results of 1:1 Verification and 1:N Identification obtained using the AdaFace model (IR-50) for LSR and other SR methods on the SCface dataset.

SR Method	SCface			
	AUC	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
LR	0.952	51.54	77.38	86.77
SR3	0.919	35.54	61.08	72.46
SPARNet	0.907	33.85	61.08	73.54
GFPGAN	0.940	42.77	67.85	79.69
SwinIR	0.881	30.31	53.54	65.23
LSR (Ours)	0.970	63.69	85.69	92.46

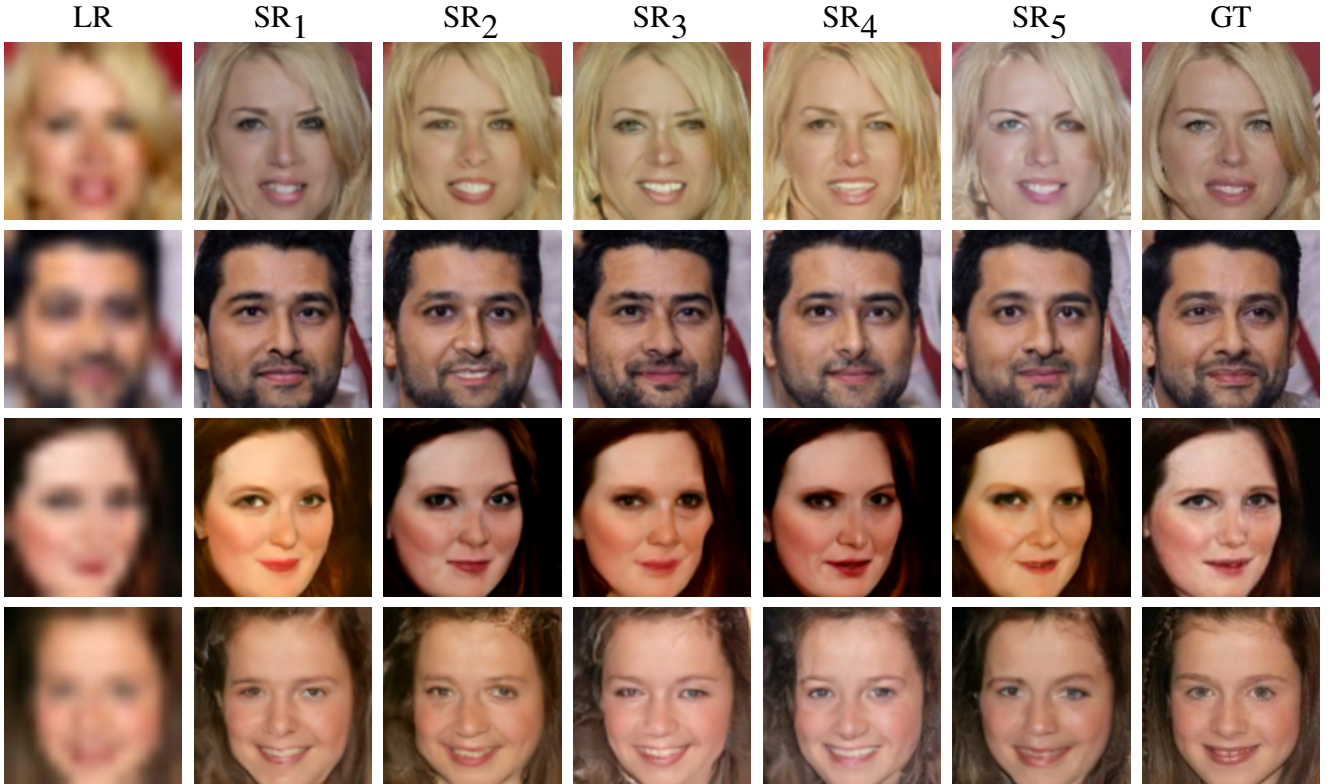


Fig. 5: **Visual results of the Lévy Diffusion Model.** The input low-resolution (LR) image is displayed alongside five generated super-resolved (SR) images ($SR_1, \dots, SR_5$) using an α -stable Lévy distribution, as well as the corresponding ground truth high-resolution (HR) image.

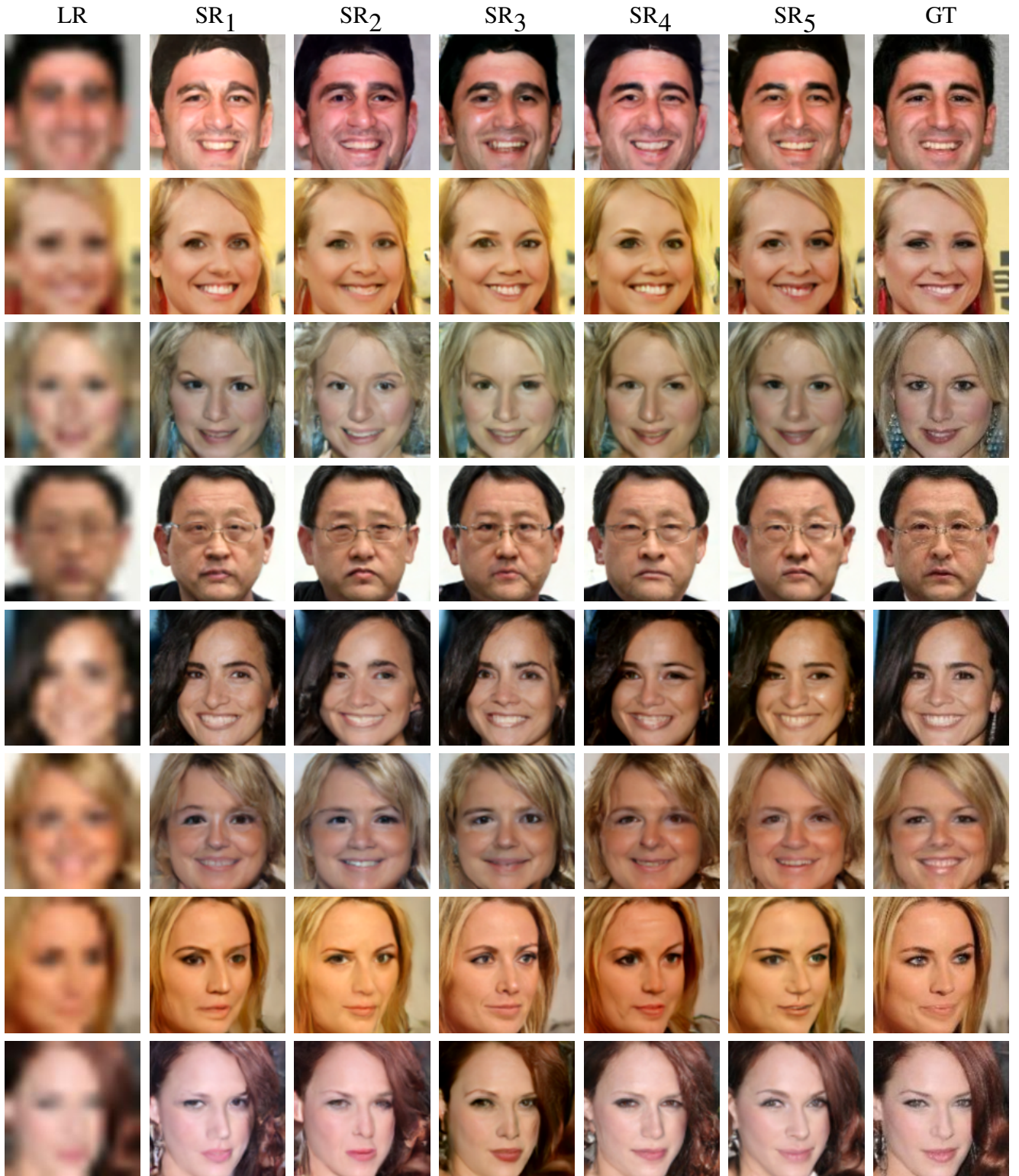


Fig. 6: **Visual results of the Lévy Diffusion Model.** The input low-resolution (LR) image is displayed alongside five generated super-resolved (SR) images (SR₁, . . . , SR₅) using an α -stable Lévy distribution, as well as the corresponding ground truth high-resolution (HR) image.