Multi-Feature Aggregation in Diffusion Models for Enhanced Face Super-Resolution

Marcelo dos Santos[1], Rayson Laroca[1,2], Rafael Ribeiro[3], João Neves[4], David Menotti[1]

[1] Federal University of Paraná, Curitiba, Brazil
[2] Pontifical Catholic University of Paraná, Curitiba, Brazil
[3] Brazilian Federal Police, Brasília, Brazil
[4] University of Beira Interior, Covilhã, Portugal

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

Contents

Introduction

- **Ill-Posed Nature of SR Problem**
  - Difficulty in recovering details (eyeglasses, beards, mustaches) and maintaining a reliable identity.

Introduction

- **Ill-Posed Nature of SR Problem**
  - Difficulty in recovering details (eyeglasses, beards, mustaches) and maintaining a reliable identity.
- **Limitations of Existing Approaches**
  - Attribute-assisted super-resolution: require classifiers or manual extraction of features (inefficient).

Introduction

- **Ill-Posed Nature of SR Problem**
  - Difficulty in recovering details (eyeglasses, beards, mustaches) and maintaining a reliable identity.
- **Limitations of Existing Approaches**
  - Attribute-assisted super-resolution: require classifiers or manual extraction of features (inefficient).
- **It is necessary to automatically capture features.**
  - Facial proportions, shapes and other more abstract features.

Proposed Method

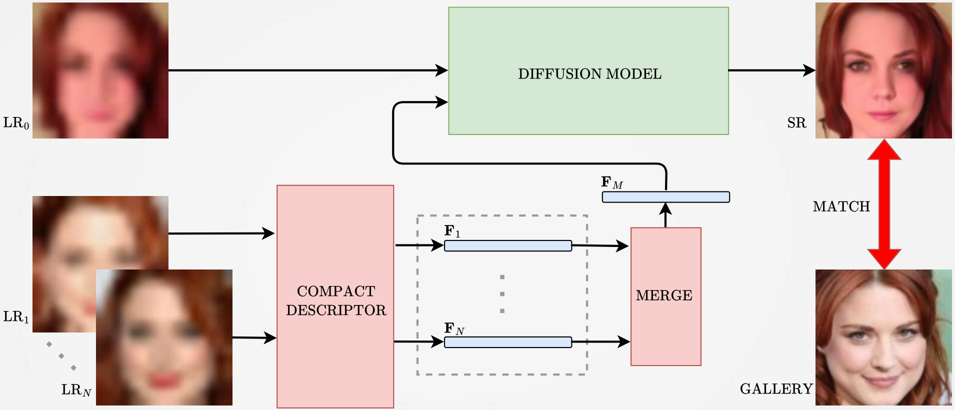

Figure: Overview of the proposed method.

Diffusion Models

Forward SDE (data $\rightarrow$ noise)

$$\mathbf{x}(0) \longrightarrow \mathrm{d}\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \longrightarrow \mathbf{x}(T)$$

$$\mathbf{x}(0) \longleftarrow \mathrm{d}\mathbf{x} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \overset{\text{score function}}{\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}}]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \longleftarrow \mathbf{x}(T)$$

Reverse SDE (noise $\rightarrow$ data)

Diffusion Models

Forward SDE (data $\rightarrow$ noise)

$$\mathbf{x}(0) \longrightarrow \mathbf{dx} = \mathbf{f}(\mathbf{x}, t)\mathrm{d}t + g(t)\mathrm{d}\mathbf{w} \longrightarrow \mathbf{x}(T)$$

$$\mathbf{x}(0) \longleftarrow \mathbf{dx} = [\mathbf{f}(\mathbf{x}, t) - g(t)^2 \underbrace{\boxed{\nabla_{\mathbf{x}} \log p_t(\mathbf{x})}}_{\text{score function}}]\mathrm{d}t + g(t)\mathrm{d}\bar{\mathbf{w}} \longleftarrow \mathbf{x}(T)$$

Reverse SDE (noise $\rightarrow$ data)

- Loss function

$$\min_{\theta} \mathbb{E}_{t\sim\mathcal{U}[0,T]} \mathbb{E}_{\mathbf{x}_0\sim p(\mathbf{x}_0)} \mathbb{E}_{\mathbf{x}(t)\sim p_t(\mathbf{x}(t)|\mathbf{x}(0))} \left[\| s_\theta(\mathbf{x}(t), \mathbf{y}, \mathbf{F}_M, t) - \nabla_{\mathbf{x}(t)} \log p(\mathbf{x}(t)|\mathbf{x}(0)) \|_2^2 \right] \tag{1}$$

Face Reconstruction

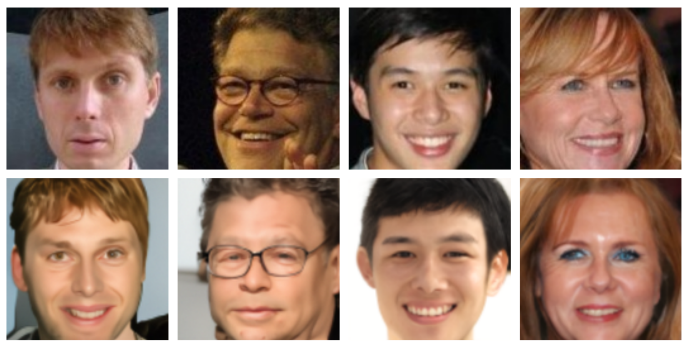

Figure: First row: original HR images. Second row: synthetic HR images.

Experimental Setup

- **Datasets**
  - **FFHQ**: Used for model training.
  - **CelebA**: Used for testing with 500 identities.
  - **Quis-Campi**: Real-world surveillance dataset with 90 identities.

Experimental Setup

- **Datasets**
  - **FFHQ**: Used for model training.
  - **CelebA**: Used for testing with 500 identities.
  - **Quis-Campi**: Real-world surveillance dataset with 90 identities.
- **Algorithm Parameters**
  - Noise levels: $\sigma_{\text{min}} = 0.001$, $\sigma_{\text{max}} = 348$.
  - LR images: $8\times$ downsampling and bicubic upsampling to 128×128.
  - SDE solved with $2,000$ steps for image reconstruction.

Experimental Setup

- **Datasets**
  - **FFHQ**: Used for model training.
  - **CelebA**: Used for testing with 500 identities.
  - **Quis-Campi**: Real-world surveillance dataset with 90 identities.
- **Algorithm Parameters**
  - Noise levels: $\sigma_{\text{min}} = 0.001$, $\sigma_{\text{max}} = 348$.
  - LR images: $8\times$ downsampling and bicubic upsampling to 128×128.
  - SDE solved with $2,000$ steps for image reconstruction.
- **Feature Extraction and Recognition**
  - 512-dimensional feature vectors using AdaFace with ResNet backbone.
  - Cosine similarity metric for image comparison.
  - Recognition by comparing SR images against gallery images.

Experimental Setup

- **Datasets**
  - **FFHQ**: Used for model training.
  - **CelebA**: Used for testing with 500 identities.
  - **Quis-Campi**: Real-world surveillance dataset with 90 identities.
- **Algorithm Parameters**
  - Noise levels: $\sigma_{\text{min}} = 0.001$, $\sigma_{\text{max}} = 348$.
  - LR images: $8\times$ downsampling and bicubic upsampling to 128×128.
  - SDE solved with $2,000$ steps for image reconstruction.
- **Feature Extraction and Recognition**
  - 512-dimensional feature vectors using AdaFace with ResNet backbone.
  - Cosine similarity metric for image comparison.
  - Recognition by comparing SR images against gallery images.
- **Comparison with State-of-the-Art**
  - Compared against diffusion models algorithms SR3, SDE-SR, IDM and SRDG.

Results

SR Method	AUC	Rank-1 (%)	Rank-5 (%)	Rank-10 (%)
LR	0.885	27.00	41.40	51.60
SR3	0.936	45.60	62.00	71.00
SDE-SR	0.933	48.60	66.60	72.40
FASR (Ours)	0.946	52.80	70.00	76.00

Tabela: The 1:1 verification and 1:N identification results obtained using the AdaFace recognition model through super-resolution on the CelebA dataset.

Results

SR Method	AUC	Rank-1(%)	Rank-5(%)	Rank-10(%)
LR	0.816	23.78	46.89	58.67
IDM	0.885	28.22	56.44	70.00
SR3	0.914	45.78	69.56	79.77
SDE-SR	0.917	50.00	72.67	81.56
SRDG	0.920	49.33	73.11	82.00
FASR (Ours)	0.917	51.33	72.44	80.00

Tabela: The 1:1 verification and 1:N identification results obtained using the AdaFace recognition model through super-resolution on the Quis-Campi dataset.

Results

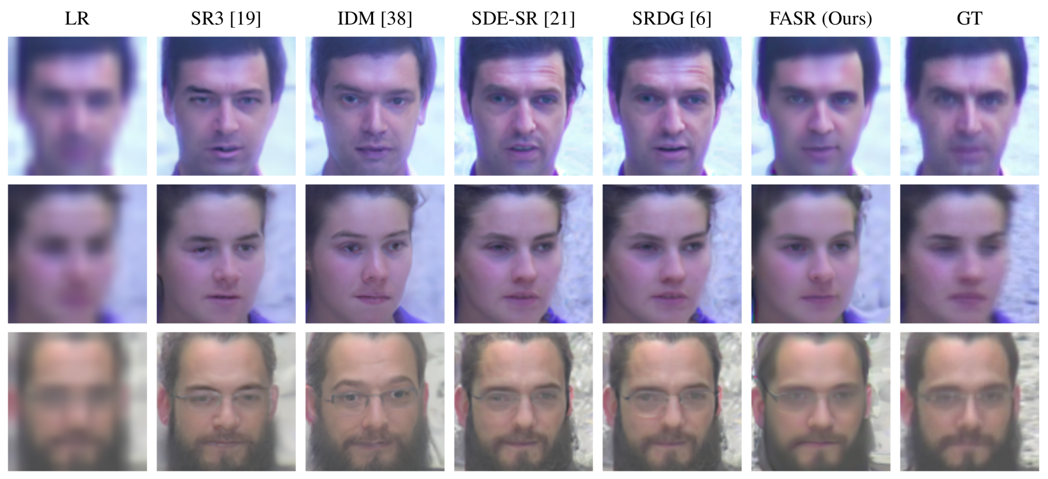

Figure: Comparison of low-resolution (LR), super-resolution (SR) and ground gruth (GT) images from the Quis-Campi dataset.

Results

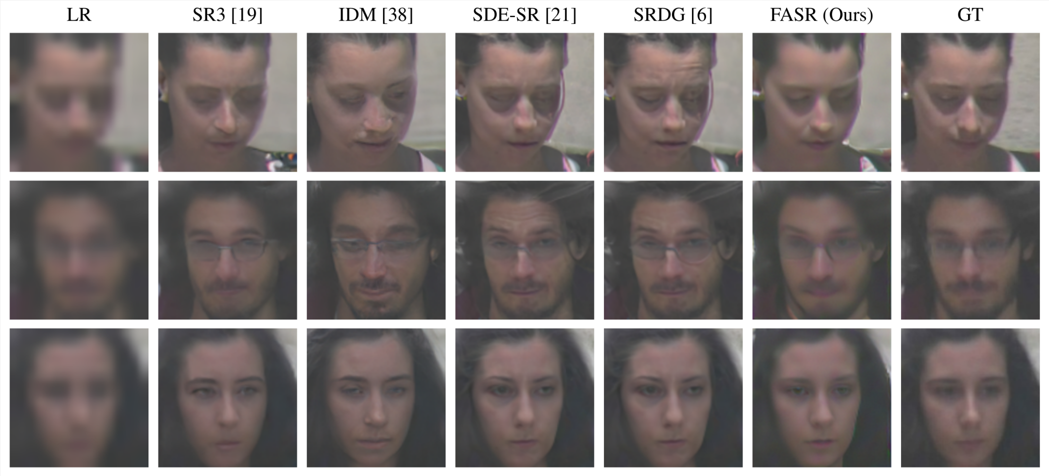

Figure: Comparison of low-resolution (LR), super-resolution (SR) and ground gruth (GT) images from the Quis-Campi dataset.

Results

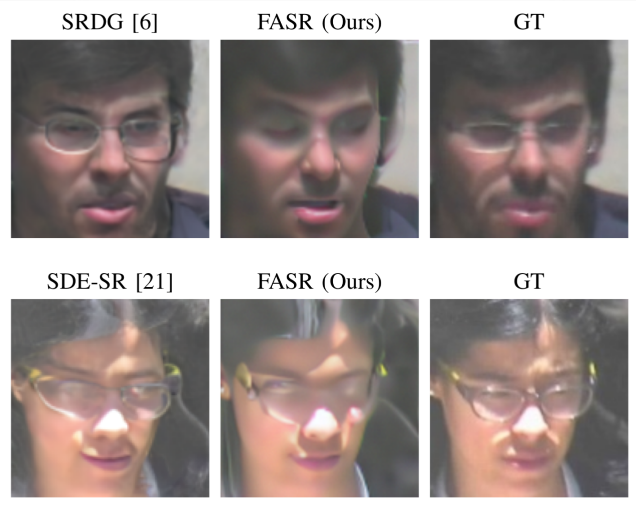

Figure: Some failure cases of the proposed approach.

Conclusions

- We introduced FASR, an algorithm that integrates multi-features and a reference LR image into diffusion models to generate SR images.

Conclusions

- We introduced FASR, an algorithm that integrates multi-features and a reference LR image into diffusion models to generate SR images.
- A key advantage of our algorithm is that it utilizes automatically extracted features.

Conclusions

- We introduced FASR, an algorithm that integrates multi-features and a reference LR image into diffusion models to generate SR images.
- A key advantage of our algorithm is that it utilizes automatically extracted features.
- Our algorithm preserves individuals' identities more effectively than other methods.

Conclusions

- We introduced FASR, an algorithm that integrates multi-features and a reference LR image into diffusion models to generate SR images.
- A key advantage of our algorithm is that it utilizes automatically extracted features.
- Our algorithm preserves individuals' identities more effectively than other methods.
- FASR produces high-quality images with enhanced face symmetry, reduced noise and minimized distortions in face attributes.

Conclusions

- We introduced FASR, an algorithm that integrates multi-features and a reference LR image into diffusion models to generate SR images.
- A key advantage of our algorithm is that it utilizes automatically extracted features.
- Our algorithm preserves individuals' identities more effectively than other methods.
- FASR produces high-quality images with enhanced face symmetry, reduced noise and minimized distortions in face attributes.
- We achieved state-of-the-art results for recognition metrics on the CelebA and Quis-Campi datasets.

Thank you!