

INTERPRETABLE FACE QUALITY ASSESEMENT

Byungho et al, 2023

Matheus Levi, 03 de Outubro de 2024

SUMÁRIO

Introdução

01

Motivação

02

Métrica
Proposta

03

Análise
Quantitativa

04

Análise
Qualitativa

05

Conclusão

06



INTRODUÇÃO

INTRODUÇÃO



AVALIAÇÃO DE RECONSTRUÇÃO DE IMAGENS

- ◆ Com o aumento de interesse em reconstrução de imagens, surge a necessidade de mensurar a qualidade da restauração
- ◆ Este artigo apresenta uma métrica específica para qualidade de reconstrução **facial**

Reference

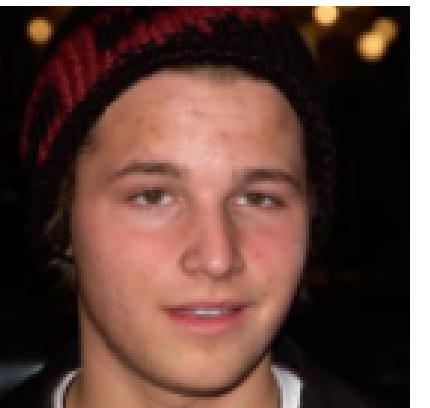


Image A

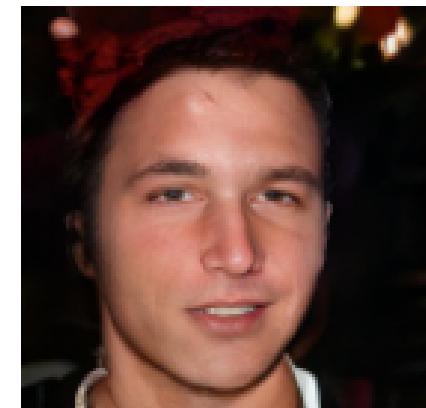


Image B

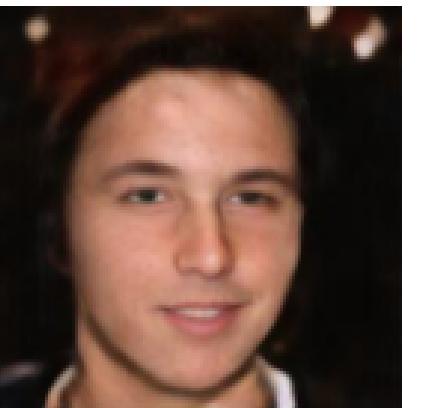


Fig. 1: Exemplos de imagens reconstruídas

MÉTRICAS EXISTENTES



PEAK SIGNAL-TO-NOISE RATIO (PSNR)

- ◆ Mensura a qualidade da imagem baseado na diferença de intensidade entre os pixels

$$PSNR = 10 \times \log \left(\frac{MAX^2}{MSE} \right)$$

MÉTRICAS EXISTENTES



PEAK SIGNAL-TO-NOISE RATIO (PSNR)

- ◆ Mensura a qualidade da imagem baseado na diferença de intensidade entre os pixels

$$PSNR = 10 \times \log \left(\frac{MAX^2}{MSE} \right)$$



STRUCTURAL SIMILARITY INDEX (SSIM)

- ◆ Avalia a similaridade entre duas imagens utilizando média e variância entre pixels.
- ◆ O valor varia de -1 a 1, sendo 1 a melhor qualidade.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

MÉTRICAS EXISTENTES



LEARNED PERCEPTUAL IMAGE PATCH SIMILARITY (LPIPS)

- ◆ Utiliza redes neurais profundas para medir a similaridade perceptual entre imagens, capturando diferenças que humanos notariam
- ◆ Rede Siamesa Convolucional gera feature maps das imagens, que são em seguida, comparadas

$$LPIPS(x, y) = \sum_{l=1}^n w_l \times d_l(f_l(x), f_l(y))$$

MÉTRICAS EXISTENTES



LEARNED PERCEPTUAL IMAGE PATCH SIMILARITY (LPIPS)

- ◆ Utiliza redes neurais profundas para medir a similaridade perceptual entre imagens, capturando diferenças que humanos notariam
- ◆ Rede Siamesa Convolucional gera feature maps das imagens, que são em seguida, comparadas

$$LPIPS(x, y) = \sum_{l=1}^n w_l \times d_l(f_l(x), f_l(y))$$



NATURAL IMAGE QUALITY EVALUATOR (NIQE)

- ◆ Baseado em estatísticas de imagens naturais e não precisa de uma imagem de **referência**
- ◆ Calcula estatísticas gaussianas na imagem e compara com um modelo de referência treinado em imagens “naturais”

MÉTRICAS EXISTENTES



BLIND/REFERENCELESS IMAGE SPATIAL QUALITY EVALUATOR (BRISQUE)

- ◆ Mede a qualidade com base em características estatísticas da imagem no domínio espacial.
- ◆ Extrai da imagem um vetor de características como média, variância, kurtosis e skewness, a partir das intensidades dos pixels.
- ◆ Esse vetor é comparado a um modelo treinado em imagens de alta qualidade, usando uma SVM

MÉTRICAS EXISTENTES



BLIND/REFERENCELESS IMAGE SPATIAL QUALITY EVALUATOR (BRISQUE)

- ◆ Mede a qualidade com base em características estatísticas da imagem no domínio espacial.
- ◆ Extrai da imagem um vetor de características como média, variância, kurtosis e skewness, a partir das intensidades dos pixels.
- ◆ Esse vetor é comparado a um modelo treinado em imagens de alta qualidade, usando uma SVM



FACE IMAGE QUALITY ASSESSMENT (FIQA)

- ◆ Baseia-se em características que influenciam o desempenho do modelo de reconhecimento facial, como iluminação, pose e nitidez.
- ◆ É utilizado como método de seleção de imagens de qualidade para treinamento de modelos de reconhecimento facial



MOTIVAÇÃO

MOTIVAÇÃO

- ◆ Muitos trabalhos recentes de restauração facial utilizaram pesquisas de opinião humana na sua análise qualitativa
- ◆ Este tipo de avaliação carece de escala e tem um custo maior para ser executado
- ◆ Estudos indicam que objetos e faces são reconhecidos por partes diferentes do cérebro
- ◆ Propor uma métrica que leve em consideração regiões faciais e se assemelhe ao julgamento humano

PROVA DE CONCEITO

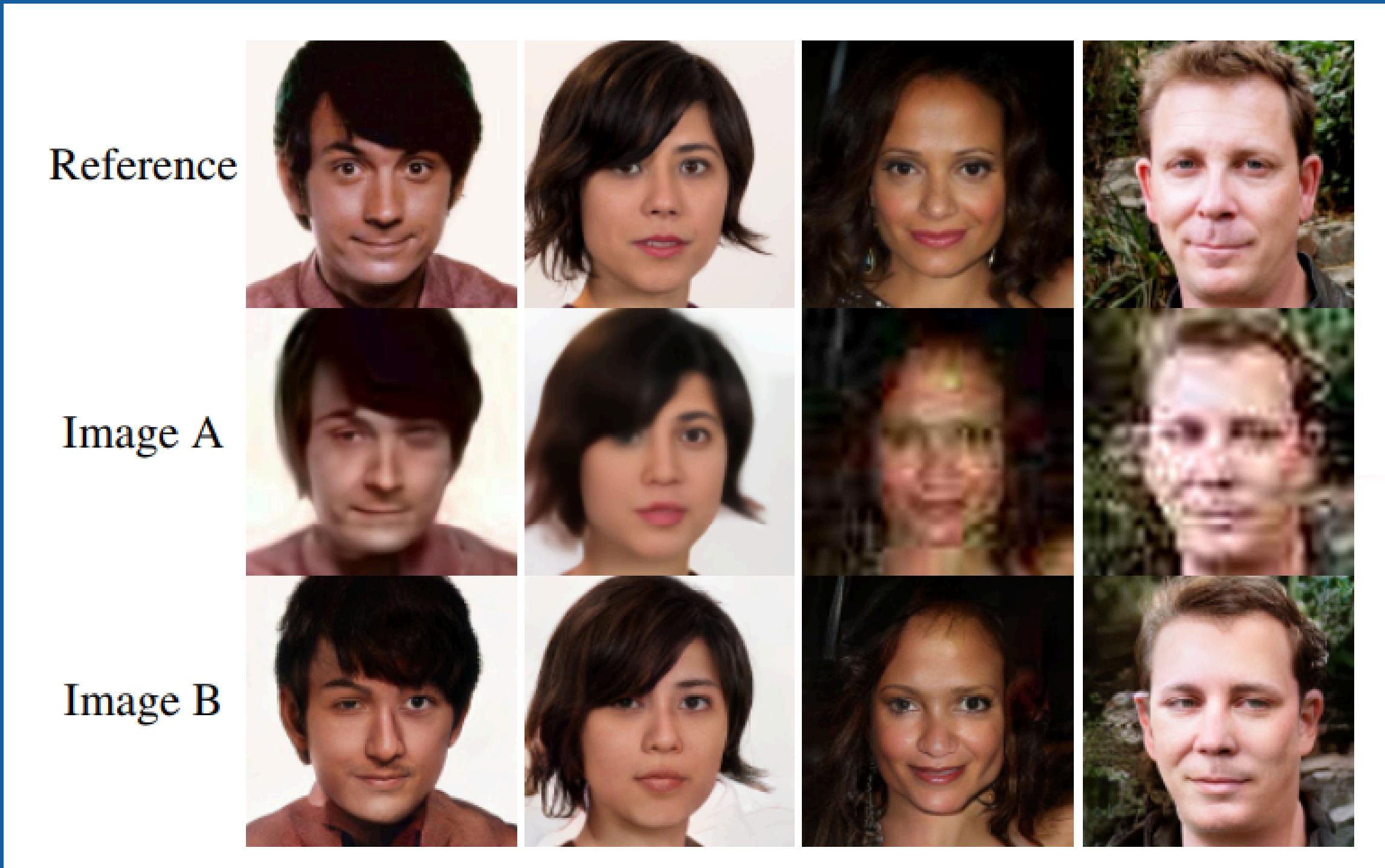


Fig. 2: Imagens utilizadas na pesquisa

PROVA DE CONCEITO

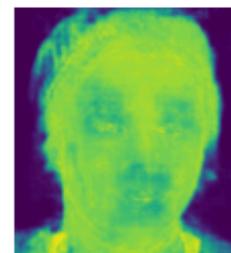
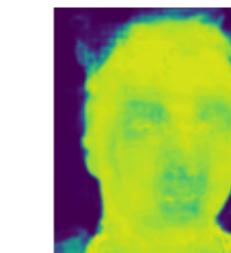
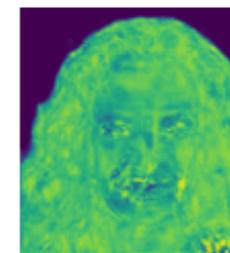
	w/ Reference		w/o Reference	
Reference	Image A	Image B	Image A	Image B
				
				
PSNR [21]		✓		N/A
SSIM [51]		✓		N/A
LPIPS [57]	✓			N/A
NIQE [39]			✓	
BRISQUE [38]	✓			✓
PI [1]	✓			✓
FIQA [16, 15, 46, 40]	✓			✓
IFQA (Ours)	✓			
Human	✓		✓	✓
Human judgment: ✓		Full-Ref. IQA: ✓		No-Ref. IQA: ✓

Fig. 3: Comparação de métricas com julgamento humano



MÉTRICA PROPOSTA

GERADOR PARA RESTAURAÇÃO

- ◆ Gerador treinado para restaurar imagens LQ em imagens HQ
256x256
- ◆ Imagens LQ são obtidas corrompendo as imagens HQ da base de dados (FFHQ) da seguinte maneira:

$$I_{LQ} = ((I_{HQ} \otimes k) \downarrow_r + n_\sigma)_{JPEG_q},$$

k = kernel gaussiano/motion-blur

r = downsampling [0.4, 0.9)

n = gaussian noise [50, 250)

q = fator de compressão do JPEG [5,50)

DISCRIMINADOR PARA AVALIAÇÃO

- ◆ Gerador treinado para avaliar a qualidade das imagens (LQ/HQ)
- ◆ Arquitetura baseada em U-net que permite julgar a imagem pixel a pixel ao invés da imagem inteira
- ◆ Gera uma saída interpretável por regiões e permite o modelo focar em certas partes da face

FACIAL PRIMARY REGION SWAP

- ◆ Técnica de aumento de dados (data augmentation) que consiste em alternar regiões faciais de uma imagem entre HQ e LQ

FACIAL PRIMARY REGION SWAP

- ◆ Técnica de aumento de dados (data augmentation) que consiste em alternar regiões faciais de uma imagem entre HQ e LQ
- ◆ Primeiro, é aplicado um modelo de detecção de landmarks para obter as regiões faciais de uma imagem em HQ

FACIAL PRIMARY REGION SWAP

- ◆ Técnica de aumento de dados (data augmentation) que consiste em alternar regiões faciais de uma imagem entre HQ e LQ
- ◆ Primeiro, é aplicado um modelo de detecção de landmarks para obter as regiões faciais de uma imagem em HQ
- ◆ Logo após, essas regiões são trocadas aleatoriamente por regiões de baixa qualidade

$$\begin{aligned} I_{HQ \rightarrow LQ/RF} &= \mathbf{M}_{\text{FPRS}} \odot I_{HQ} \\ &+ (1 - \mathbf{M}_{\text{FPRS}}) \odot I_{LQ/RF} \end{aligned}$$

SUPERVISÃO DO MODELO

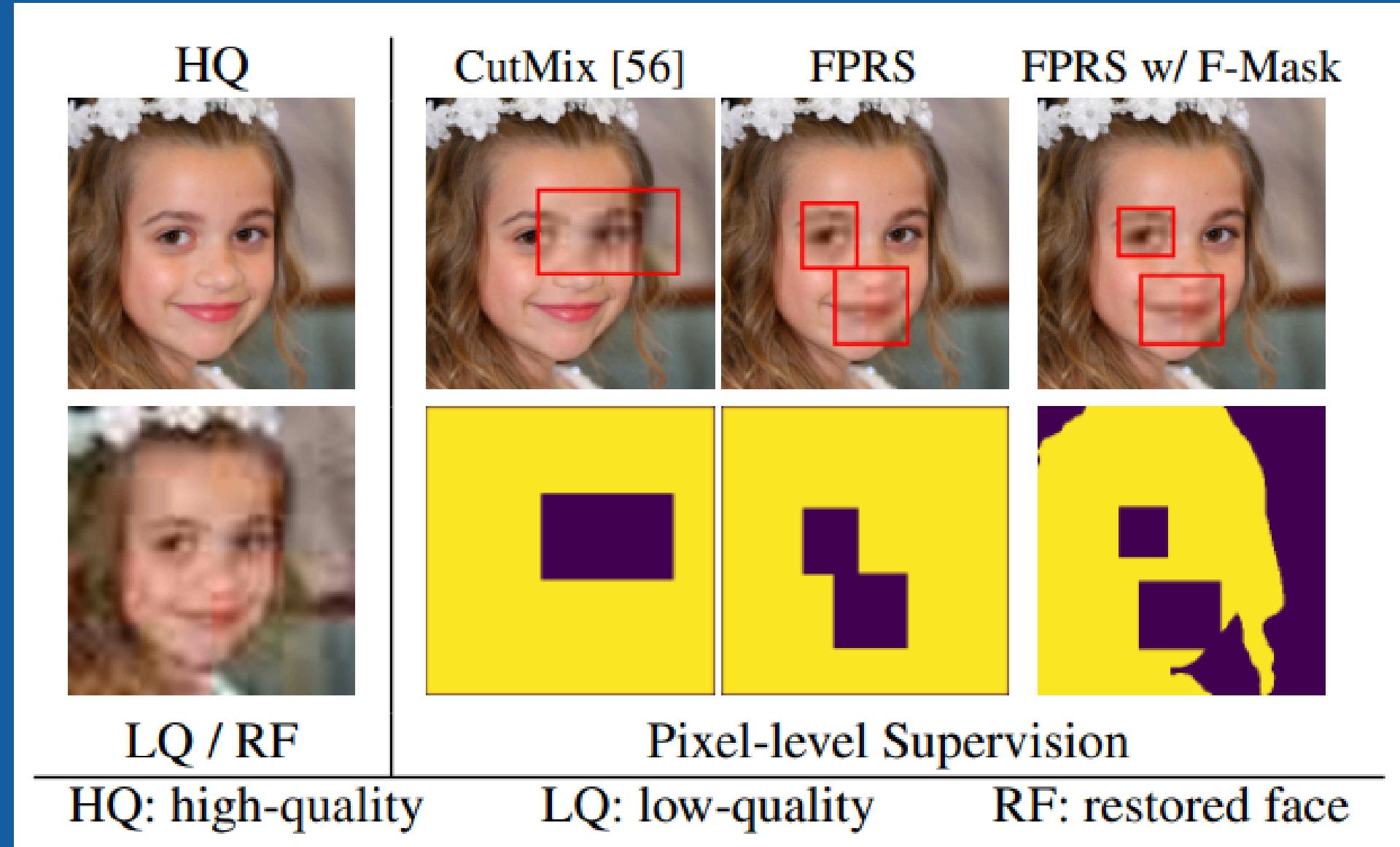


Fig. 4: Exemplo de Supervisão do modelo

OUTRAS INFORMAÇÕES DO MODELO

PERDA ADVERSARIAL

- ◆ Utiliza o objetivo Least Square (LSGAN), função mais comum em GANs atualmente, o clássico gerador treinado para “enganar” o discriminador

$$\mathcal{L}_{adv,G} = \mathbb{E}_{I_{RF}} [(D^U(I_{RF}) - \mathbf{1}))^2],$$

OUTRAS INFORMAÇÕES DO MODELO

PERDA ADVERSARIAL

- ◆ Utiliza o objetivo Least Square (LSGAN), função mais comum em GANs atualmente, o clássico gerador treinado para “enganar” o discriminador

$$\mathcal{L}_{adv,G} = \mathbb{E}_{I_{RF}} [(D^U(I_{RF}) - 1))^2],$$

PERDA PIX

- ◆ Objetivo é forçar o gerador a criar imagens similares a imagem HQ correspondente, comparando-as pixel a pixel

$$\mathcal{L}_{pix} = \mathbb{E}_{I_{RF}, I_{HQ}} [||I_{RF} - I_{HQ}||_2].$$

OUTRAS INFORMAÇÕES DO MODELO

PERDA ADVERSARIAL

- ◆ Versão do discriminador da perda adversarial, em que \mathbf{M}_{FACE} é uma máscara binária que possui 1s na região da face

$$\mathcal{L}_{adv,D} = \mathbb{E}_{I_{HQ}}[(D^U(I_{HQ}) - \mathbf{M}_{\text{FACE}})^2] + \mathbb{E}_{I_{LQ}, I_{RF}}[D^U(I_{LQ/RF})^2],$$

OUTRAS INFORMAÇÕES DO MODELO

PERDA ADVERSARIAL

- ◆ Versão do discriminador da perda adversarial, em que \mathbf{M}_{FACE} é uma máscara binária que possui 1s na região da face

$$\mathcal{L}_{adv,D} = \mathbb{E}_{I_{HQ}}[(D^U(I_{HQ}) - \mathbf{M}_{\text{FACE}})^2] + \mathbb{E}_{I_{LQ}, I_{RF}}[D^U(I_{LQ/RF})^2],$$

PERDA PERCEPTIVA

- ◆ Para ajudar a produzir imagens fotorrealistas, é utilizada uma perda que compara as features das imagens utilizando uma rede pré-treinada VGG-19

$$\mathcal{L}_{vgg} = \sum_i ||f_i(I_{RF}) - f_i(I_{HQ})||_1,$$

FUNÇÃO OBJETIVO

$$\min_G \max_D \quad \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{pix} + \lambda_2 \mathcal{L}_{vgg},$$

$$\lambda_1 = 50$$

$$\lambda_2 = 5$$

FRAMEWORK PROPOSTO

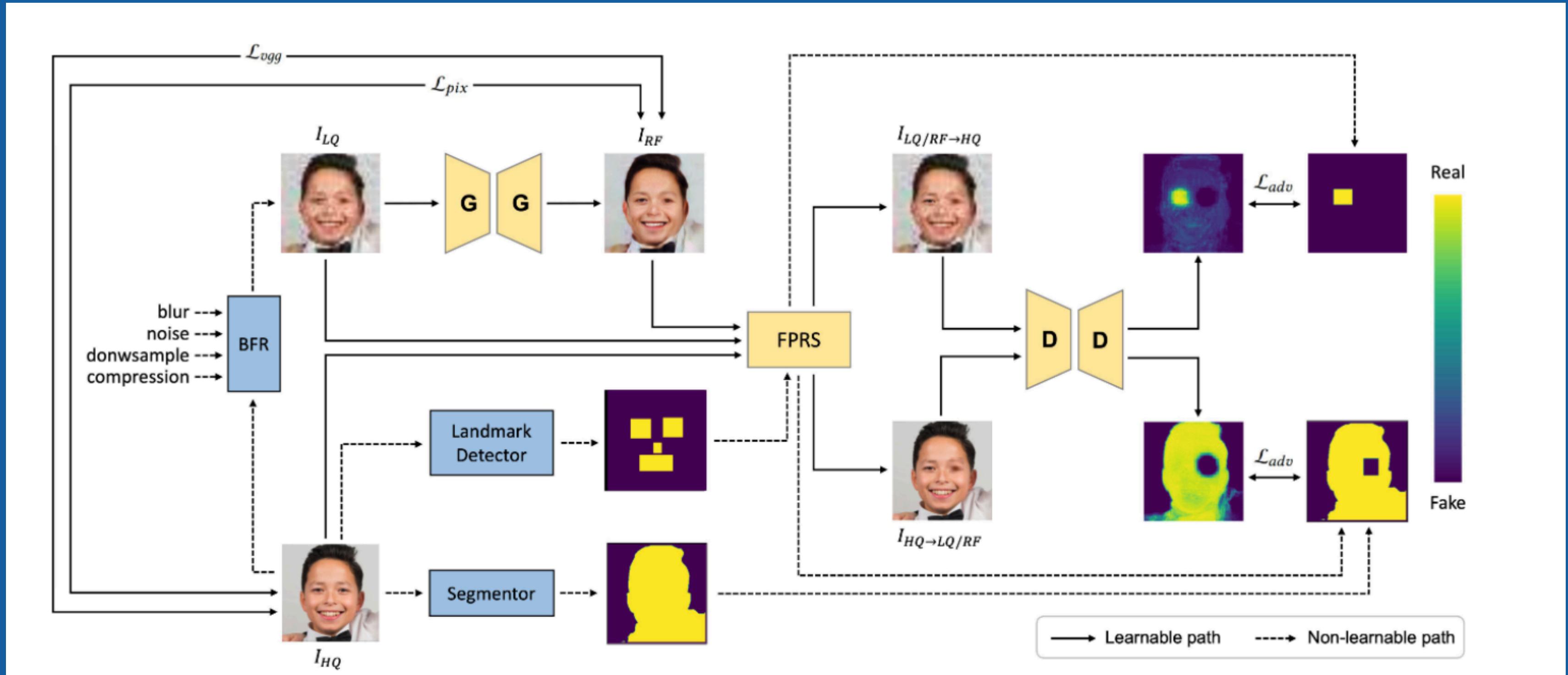


Fig. 5: Arquitetura proposta



ANÁLISE QUANTITATIVA

AVALIAÇÃO HUMANA

Participantes ranquearam as imagens de acordo com o realismo

- ◆ Participantes receberam pacotes de 6 imagens, uma imagem LQ e 5 imagens geradas por modelos de restauração
- ◆ O número total de imagens utilizadas nesse estudo foi de 1200, com 6000 avaliações no total

AVALIAÇÃO HUMANA

Participantes ranquearam as imagens de acordo com o realismo

- ◆ Participantes receberam pacotes de 6 imagens, uma imagem LQ e 5 imagens geradas por modelos de restauração
- ◆ O número total de imagens utilizadas nesse estudo foi de 1200, com 6000 avaliações no total
- ◆ Foram utilizadas medidas como Spearman's rank order correlation coefficient (SRCC) e Kendall rank order correlation coefficients (KRCC), que são utilizadas para comparar julgamento humano com outras métricas

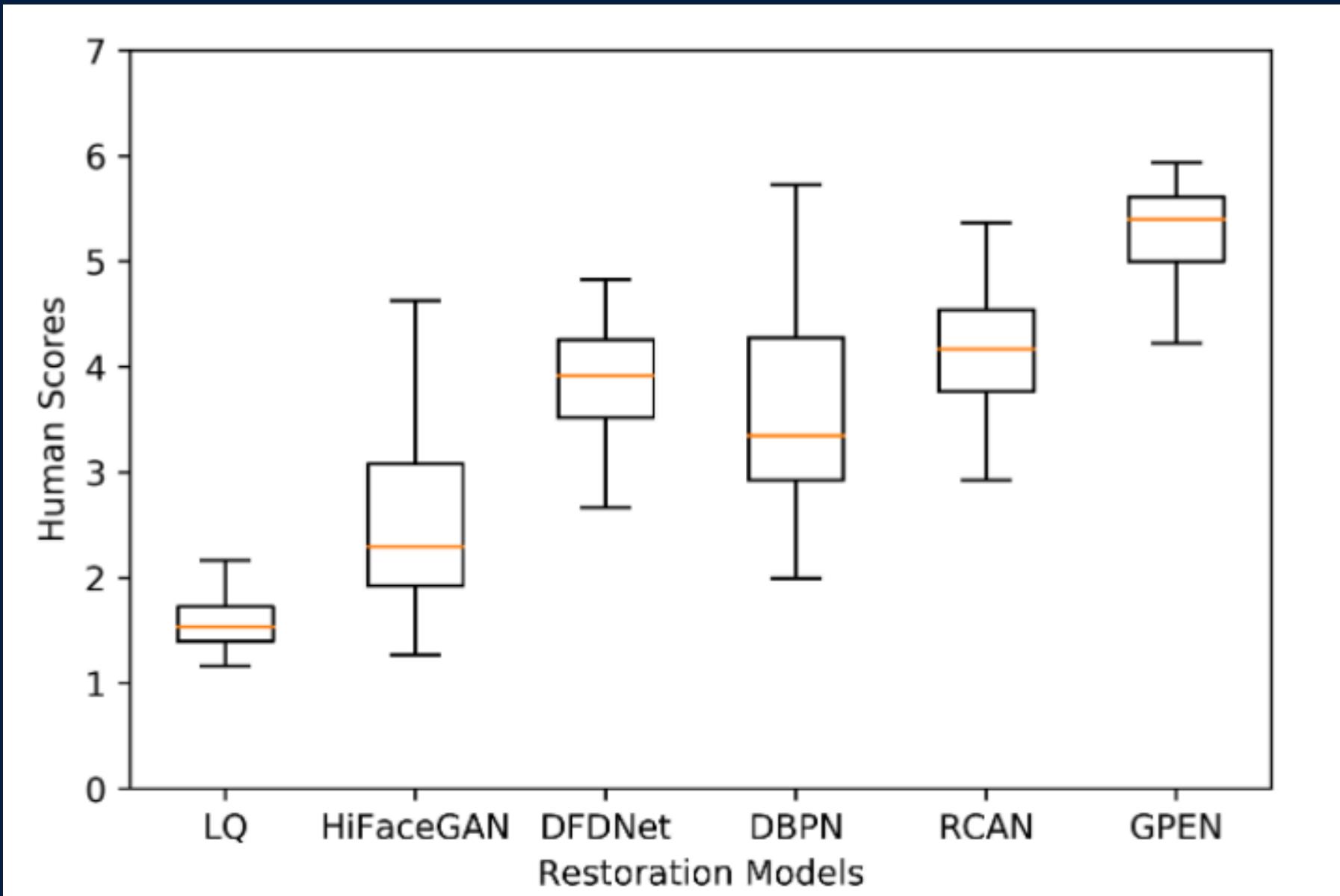


Fig. 6: Box plot dos modelos de restauração

ANÁLISE COMPARATIVA

Comparando as métricas com o julgamento humano

Metric	Type	SRCC ↑	KRCC ↑
NIQE [39]		0.2668	0.2039
PI [1]	NR-IQA	0.4125	0.3173
BRISQUE [38]	(General)	0.4405	0.3373
IFQA (Ours)		0.6400	0.5186
SER-FIQ [46]		0.3554	0.2706
FaceQnet-V1 [15]	NR-IQA (FIQA)	0.4560	0.3453
FaceQnet-V0 [16]		0.5491	0.434
SDD-FIQA [40]		0.5920	0.4840

Tab. 1: Análise comparativa no FFHQ e CelebA-HQ

ANÁLISE COMPARATIVA

Comparando as métricas com o julgamento humano

Metric	Type	SRCC ↑	KRCC ↑
NIQE [39]		0.5005	0.4053
PI [1]	NR-IQA	0.6382	0.5320
BRISQUE [38]	(General)	0.6451	0.5573
IFQA (Ours)		0.6988	0.6013
SER-FIQ [46]		0.1657	0.1386
FaceQnet-V1 [15]	NR-IQA (FIQA)	0.2725	0.2106
FaceQnet-V0 [16]		0.4474	0.3813
SDD-FIQA [40]		0.5131	0.4120

Tab. 2: Análise comparativa no IWF



ANÁLISE QUALITATIVA

COMPARAÇÃO VISUAL

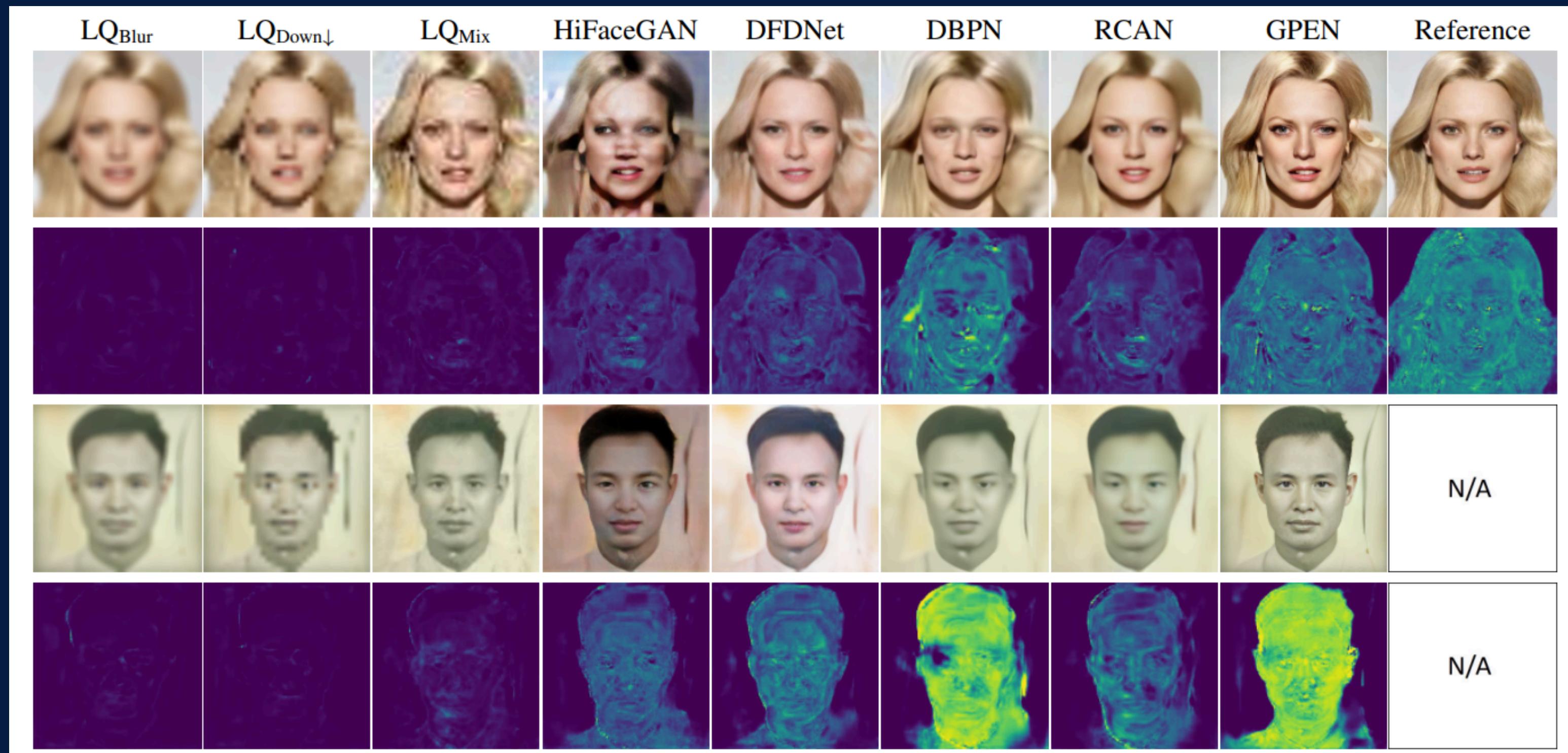


Fig. 6: Visualização Interpretável de IFQA

COMPARAÇÃO VISUAL

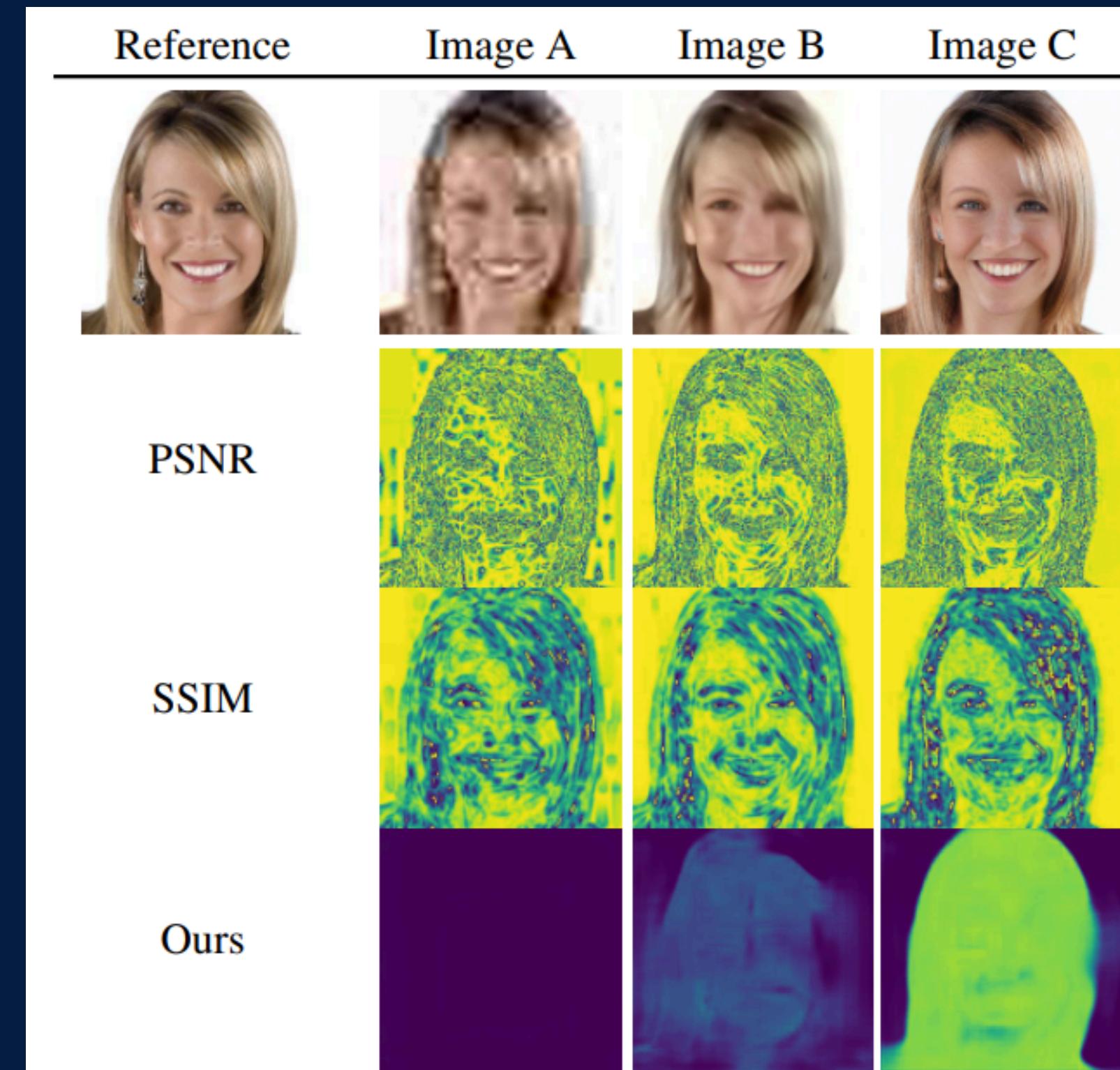


Fig. 7: Comparação visual com outras métricas pixel a pixel

ABLAÇÃO

Discriminator	SRCC ↑	KRCC ↑
Baseline (single-output)	0.5885	0.4840
Baseline (per-pixel)	0.4674	0.3820
Baseline (per-pixel) + CutMix [56]	0.5437	0.4420
Baseline (per-pixel) + FPRS	0.6265	0.5213
Baseline (per-pixel) + FPRS + F-Mask	0.6694	0.5600

F-Mask: facial masks using a segmentation model

Tab. 3: Ablação feita com os componentes do modelo

ABLAÇÃO

Generator	Task	Parameters	SRCC ↑	KRCC ↑
GPEN [53]	FIR	pre-trained	0.4997	0.4166
DFDNet [32]	FIR		0.5391	0.4366
DBPN [12]	GIR		0.5582	0.4586
RCAN [58]	GIR		0.5711	0.4680
RCAN	GIR	learnable	0.6454	0.5480
Plain model	FIR	learnable	0.6694	0.5600

FIR: face image restoration GIR: general image restoration

Tab. 4: Comparaçāo de performance dos modelos e do gerador

ABLAÇÃO

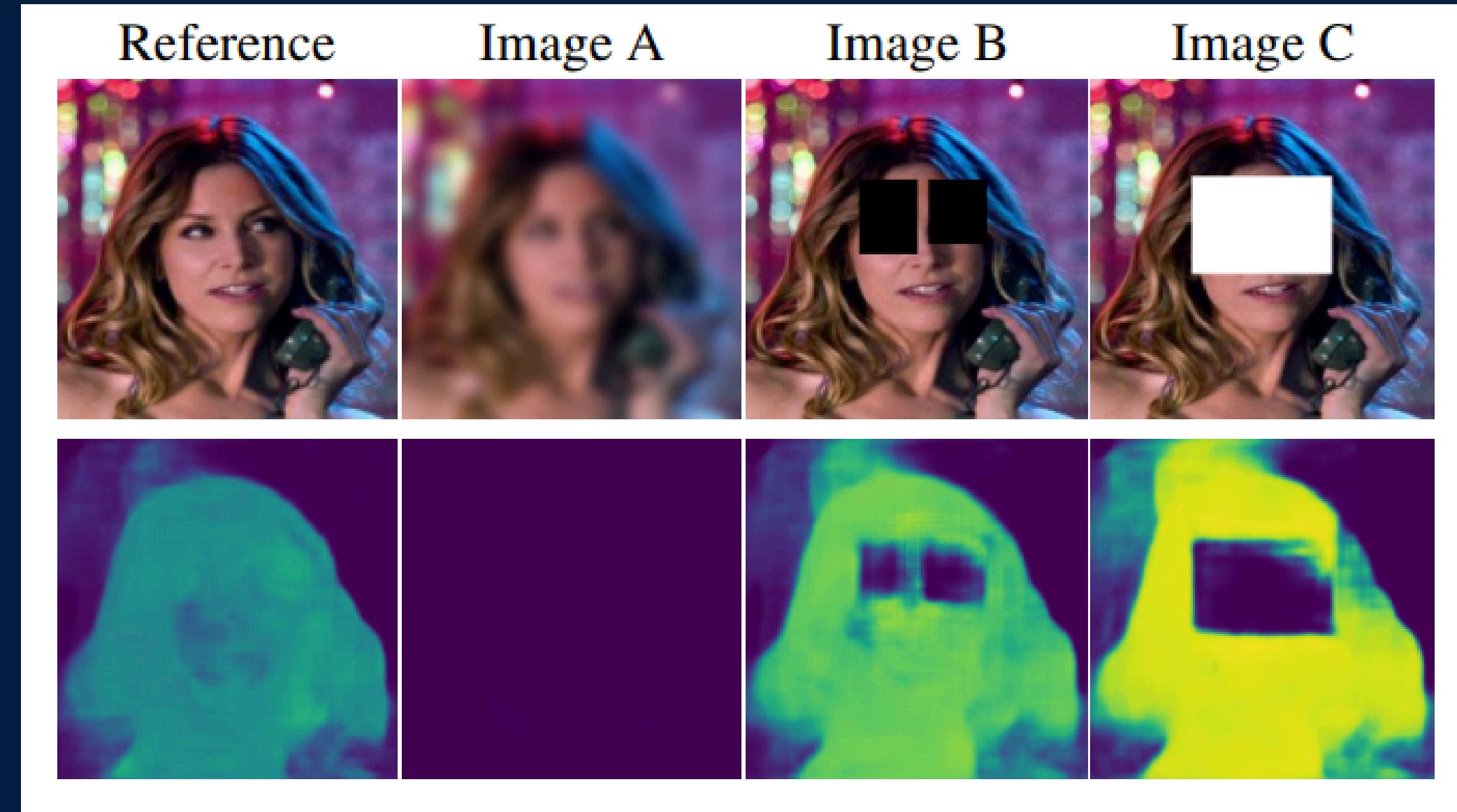


Fig. 8: IFQA em alguns cenários diferentes

ABLAÇÃO

Imagens foram manipuladas utilizando StarGAN v2

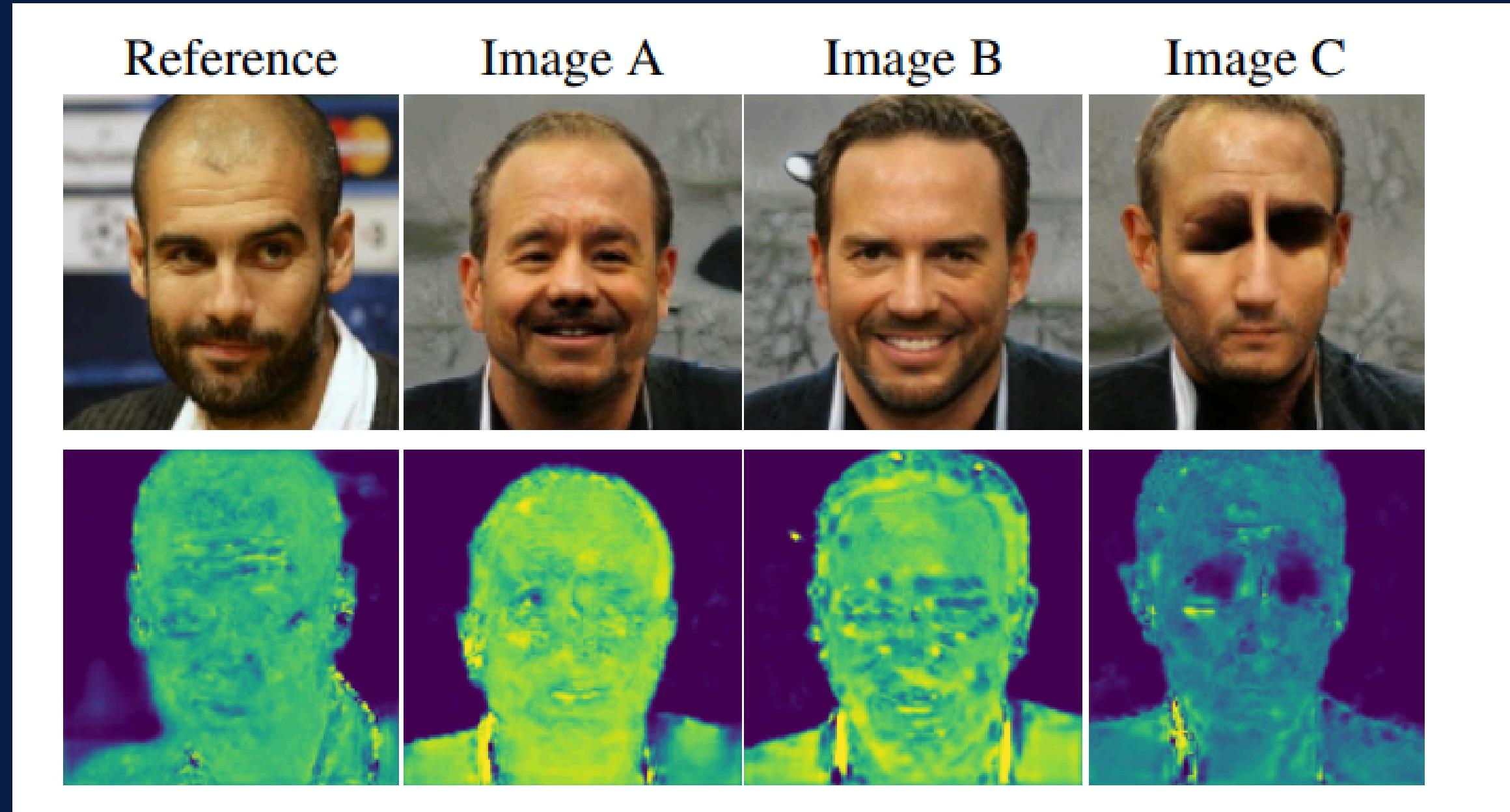


Fig. 9: IFQA em cenários de manipulação facial

ABLAÇÃO

Method	FID↓	
	Latent-guided	Reference-guided
StarGAN v2 w/o IFQA	14.6657	23.7138
StarGAN v2 w IFQA	13.8008	22.6457

Tab. 5: IFQA usado como adição na função objetivo

CONCLUSÃO

- ◆ Imagens faciais são uma categoria de imagem única, que portanto deve ser tratada com especificidade
- ◆ Distorções em certas regiões do rosto, mesmo que pequenas, impactam significativamente na percepção humana
- ◆ A métrica proposta se mostra melhor correlacionada com a percepção humana do que as métricas tradicionais

Obrigado

Referência:

<https://arxiv.org/abs/2211.07077>