Avaliação do Impacto da Super-Resolução no Reconhecimento Facial em Imagens de Baixa Resolução

Artur Vinícius Santos de Lima[1], **João Pedro Ferreira da Cruz**[1], **Leonardo Nogueira Matos**[1]

[1] Departamento de Computação – Universidade Federal de Sergipe (UFS)

{artur.delima, pedro.dacruz, leonardo}@dcomp.ufs.br

Abstract. *In facial recognition, one of the primary obstacles is the low pixel count on the face for detection and recognition. This paper investigates the impact of Super-Resolution techniques on the performance of facial recognition systems motivated by automatic classroom attendance tracking, utilizing technologies such as RRDBNet for Super-Resolução, and Deep Learning for face detection, cropping, and alignment. Using LFW as the main dataset, generating results that show a disconnection between the perceived visual and facial recognition improvement.*

Resumo. *No Reconhecimento Facial, um dos principais obstáculos é a quantidade baixa de pixels no rosto para a detecção e o reconhecimento facial. Este artigo investiga o impacto de técnicas de Super-Resolução no desempenho de sistemas de reconhecimento facial motivado pelo problema do controle automático de presença em salas de aula através de tecnologias como a RRDBNet para a Super-Resolução, Aprendizado Profundo para a detecção, recorte e alinhamento facial. Tendo como principal base o LFW e o pipeline gerou resultados que demonstram a issociação entre as métricas de reconstrução e o desempenho biométrico.*

1. Introdução

Sistemas de reconhecimento facial têm sido empregados em diferentes contextos de autenticação, monitoramento e organização de ambientes. Em salas de aula, uma aplicação possível é o controle automático de presença, a partir da identificação dos alunos através de fotos intercaladas da sala.

Entretanto, esse cenário impõe desafios relevantes em fotos tiradas em ambientes não controlados: os rostos podem aparecer distantes da câmera, parcialmente desfocados, comprimidos ou afetados por iluminação inadequada [Huang et al. 2007].

Durante anos, esses problemas impactaram modelos de reconhecimento e detecção facial, e, com o passar do tempo, diversas ferramentas para lidar com eles foram criadas, para o desenvolvimento de um pipeline completo que permitisse esse tipo de tarefa, com a descoberta dos modelos de terceira geração de reconhecimento facial advindos da origem do Aprendizado Profundo [Adjabi et al. 2020] ou de descobertas como as Redes Neurais Convolucionais, e a criação de ferramentas de Super-Resolução (SR) e outros.

Nesse contexto, tais técnicas podem ser utilizadas como uma das etapas de pré-processamento para aumentar a resolução aparente das faces antes da detecção, alinhamento e extração de características [Adjabi et al. 2020].

Destarte, este trabalho propõe analisar a integração de técnicas de SR com modelos de Aprendizado Profundo. Estes métodos, caracterizados por [Goodfellow et al. 2016] como o estado da arte em aprendizado supervisionado, consolidaram-se na área de reconhecimento facial. Modelos como o DeepFace, desenvolvido para mitigar as limitações impostas por ambientes não controlados. Para avaliar a abordagem proposta, foram realizados experimentos utilizando o conjunto de dados *Labeled Faces in the Wild* (LFW), e um *dataset* próprio criado para testar o modelo em ambientes de execução real de sala de aula e o *Celebrity Faces*.

2. Fundamentação Teórica

2.1. Detecção Facial

A detecção facial é a etapa responsável por localizar regiões de uma imagem que contém faces humanas. Diferentemente do reconhecimento facial, cujo objetivo é determinar a identidade de uma pessoa, a detecção responde a uma pergunta anterior: onde existem faces na imagem? Em um sistema de reconhecimento, essa etapa se torna importante porque define quais regiões serão recortadas, alinhadas e enviadas ao extrator de *embeddings*. Assim, erros de detecção podem afetar todo o restante do *pipeline*, mesmo quando o modelo de reconhecimento é robusto.

Formalmente, dado um quadro de entrada I, um detector facial estima um conjunto de caixas delimitadoras:

$$D(I) = \{(b_i, c_i)\}_{i=1}^{N} \tag{1}$$

em que $b_i = (x_{1,i}, y_{1,i}, x_{2,i}, y_{2,i})$ representa a caixa da i-ésima face é c_i e a confiança associada a essa detecção.

Nesse trabalho para detecção facial utilizamos o YOLO-face. Dada uma imagem de entrada, o modelo retorna um conjunto de caixas candidatas com suas respectivas confianças. O pipeline deste trabalho utiliza essas detecções como ponto de partida para o reconhecimento. A escolha baseia-se por três aspectos principais: suporte a múltiplas faces por imagem, boa relação entre custo computacional e qualidade de detecção, e facilidade de parametrização por limiar de confiança, tamanho de entrada e número de detecções.

No fluxo experimental, a consulta LR é inicialmente ampliada por interpolação ou processada por um dos modelos de Super-Resolução. Em seguida, o YOLO-face é aplicado à imagem resultante para localizar as faces. Para cada detecção aceita, a caixa retangular é convertida em uma região quadrada e recortada.

O recorte é então enviado ao MTCNN, utilizado para estimar cinco *landmarks* faciais e realizar o alinhamento. A face alinhada é finalmente processada pelo reconhecedor do InsightFace para extração do *embedding*. Essa ordem é mantida para todos os métodos comparados.

2.2. Alinhamento de rostos

O MTCNN é uma abordagem baseada em redes convolucionais em cascata para detecção e alinhamento facial [Zhang et al. 2016]. Diferentemente de detectores clássicos baseados em características manuais, ele aprende automaticamente representações visuais a partir dos dados. Sua arquitetura é composta por três estágios principais: P-Net, responsável por

gerar propostas de faces; R-Net, que refina e filtra essas propostas; e O-Net, que produz a detecção final e estima pontos faciais, como os centros dos olhos, nariz e cantos da boca.

Tendo a sua principal importância nesse trabalho a etapa de alinhamento fornecendo *landmarks* faciais e a partir deles permitindo estimar uma transformação geométrica que centraliza e normaliza a face antes da extração de *embeddings*. Dessa forma, o MTCNN atua como uma ponte entre a detecção facial e a futura etapa de reconhecimento.

2.3. Super-Resolução

Super-Resolução (SR) é uma área de processamento de imagens cujo objetivo é reconstruir uma imagem de alta resolução a partir de uma observação em baixa resolução. Em reconhecimento facial, esse problema é particularmente importante porque a perda de detalhes afeta não apenas a aparência visual da face, mas também a detecção, o alinhamento e a extração de *embeddings* utilizados por modelos de reconhecimento.

Neste trabalho, o problema é formulado como uma tarefa de SR com fator de escala $4\times$, seguindo a linha de modelos modernos de SR perceptual e generativa [Ledig et al. 2017, Wang et al. 2018]. Dada uma imagem de baixa resolução I_{LR}, o modelo aprende uma função G_θ capaz de produzir uma estimativa super-resolvida. Estimativa essa que pode ser avaliada por métricas de fidelidade, como PSNR, SSIM e MS-SSI aninhadas com a taxa de sucesso no seu uso posterior em reconhecimento e detecção facial, sucesso no alinhamento e preservação da identidade nos *embeddings*.

O modelo principal utilizado neste projeto é baseado na arquitetura RDBNet, associada ao ESRGAN [Wang et al. 2018]. Essa arquitetura utiliza blocos residuais densos sem normalização por *batch*, reduzindo artefatos e favorecendo reconstruções perceptualmente mais ricas. No contexto deste trabalho, a RRDBNet é usada como base para três estágios: um modelo inicial treinado com perda de reconstrução, um ajuste fino com perda perceptual VGG e um ajuste fino adicional com termo de preservação de identidade.

O treinamento utiliza uma combinação de funções de perda conforme o estágio experimental. A perda de reconstrução base é uma perda Charbonnier, uma versão suavizada da norma L_1:

$$\mathcal{L}_{charb}(x, y) = \sqrt{(x - y)^2 + \epsilon^2} \tag{2}$$

Para melhorar características perceptuais, é utilizada uma perda baseada em ativações de uma rede VGG pré-treinada [Johnson et al. 2016]:

$$\mathcal{L}_{vgg} = \|\phi(I_{SR}) - \phi(I_{HR})\|_1 \tag{3}$$

em que ϕ representa mapas de características extraídos por camadas intermediárias da VGG e I a imagem - sendo SR para ela super-resolvida e HR para ela em alta resolução real, chamada também de *ground-truth*. Para aproximar a reconstrução do objetivo final de reconhecimento facial, também é utilizado um termo de identidade baseado em *embeddings* extraídos por FaceNet:

$$\mathcal{L}_{id} = (1 - cos(e_{SR}, e_{HR})) + 0.1\|e_{SR} - e_{HR}\|_2 \tag{4}$$

em que e_{SR} e e_{HR} são *embeddings* das faces super-resolvida e de referência. O melhor modelo atual utiliza uma combinação balanceada de perda de reconstrução, perceptual e de identidade, com uso de *embeddings* para verificação [Schroff et al. 2015].

2.4. Reconhecimento Facial e Função de Perda

Em primeiro momento, é possível denotar a noção da complexidade nessa área pela grande quantidade de rostos e variedades. Assim, esta pode ser descrita como a busca por formas de classificar rostos e suas características. Em síntese, podendo ser dividida em 3 grandes gerações que datam de 1964 até a atualidade, sendo divididas pela abordagem de classificação [Adjabi et al. 2020] .

A partir disso, o cientista da computação Woodrow Bledsoe [Adjabi et al. 2020], utilizou um padrão que consistia no método geométrico, com extração manual da representação características como a distância entre os olhos, largura da boca, face e etc. A partir disso, foi considerada primeira geração de Reconhecimento Facial (RF).

Após certo desenvolvimento, os modelos seguiram uma abordagem mais estatística como proposto por Turk e Pentland em 1991 (Eigenfaces) [Turk and Pentland 1991], em que uma Análise de Componentes Principais, ocorre para representar o rosto como um autovetor. Por fim, o atual uso de modelos de reconhecimento facial é marcada pela criação de redes neurais. Assim modelos como o DeepFace, que utilizava de mecanismos como o alinhamento 3d da face a partir da deformação da imagem original para um modelo 3d genérico para atenuar possíveis adversidades do ambiente não controlado [Taigman et al. 2014] foram capazes de alcançar resultados satisfatórios em *datasets* como o LFW.

Nessa seção, será abordado o método de RF com uso de Redes Profundas utilizado ResNet [He et al. 2015], onde a construção utilizada são as Redes Convolucionais, que possui uma camada responsável pela correlação de informações próximas com uma operação semelhante a de convolução, e pela camada de *pooling* que diminui a complexidade computacional adquiridas na parte convolucional, preservando somente as informações relevantes adquiridas [Goodfellow et al. 2016].

O modelo ResNet [He et al. 2015] é um modelo baseado em blocos residuais, que utilizam conexões de atalho (*skip connections*) para mitigar o problema do desaparecimento do gradiente, permitindo que o sinal seja propagado de forma mais eficiente durante o processo de retropropagação e atualização dos pesos da rede. Além disso, sendo um modelo baseado em CNN, possuindo as camadas já citadas.

$$L = -\frac{1}{N}\sum_{i=1}^{N}\log\left(\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{s\cos(\theta_{y_i}+m)}+\sum_{j\neq y_i}e^{s\cos(\theta_j)}}\right) \tag{5}$$

2.5. Padrão ONNX e InsightFace

Durante o reconhecimento facial foram utilizados modelos pré-treinados na forma de ONNX, ou Open Neural Network Exchange[ONNX Community 2019]. O ONNX é um padrão aberto para representação de modelos de aprendizado de máquina, permitindo que um modelo treinado em um determinado framework seja exportado e executado em diferentes frameworks e mecanismos de inferência compatíveis, como Py-

Torch, TensorFlow, scikit-learn e XGBoost, sem a necessidade de reimplementação [ONNX Community 2019].

Para a extração de características faciais foi utilizado o modelo Buffalo_L, treinado e disponibilizado pela biblioteca InsightFace. O Buffalo_L consiste em um conjunto de modelos pré-treinados para detecção de faces, localização de pontos faciais (landmarks) e reconhecimento facial. Sendo utilizado somente o reconhecimento facial na nossa pipeline, em que possui backbone ResNet50 treinado utilizando a função de perda ArcFace no conjunto de dados WebFace600K[InsightFace 2026].

2.6. Ética em reconhecimento facial e *datasets*

Na construção do *dataset* e na validação da *pipeline*, os princípios éticos relacionados ao uso de imagens e dados faciais foram considerados durante todo o processo. Esses princípios tiveram como objetivo garantir a privacidade, a segurança e o consentimento das pessoas cujas imagens foram utilizadas. Assim, uma das medidas adotadas foi a obtenção do consentimento de todos por meio de formulário específico. Esse procedimento assegurou que todas as imagens utilizadas possuíam a autorização explícita. Somente após a confirmação desse consentimento as imagens foram incorporadas ao estudo.

Além disso, foram implementadas medidas para proteger os dados sensíveis, evitando o acesso não autorizado, o vazamento ou a divulgação indevida das imagens faciais dos participantes. Assim, buscou-se garantir que essas informações fossem utilizadas exclusivamente para os fins da pesquisa e em conformidade com os princípios éticos aplicáveis ao tratamento de dados pessoais.

3. Metodologia

A avaliação quantitativa utiliza o LFW *deepfunneled*. Os pares oficiais de *DevTrain* são usados exclusivamente como validação, enquanto os pares oficiais de *DevTest* permanecem como partição DevTest da View 1. A verificação automática do protocolo não encontrou identidades compartilhadas entre essas partições. O conjunto de teste não é utilizado para escolher checkpoint, método, hiperparâmetro ou limiar de decisão.

Em cada par de verificação, a imagem esquerda é processada como referência HR e a imagem direita como consulta. A consulta é submetida a um recorte central e a uma degradação sintética que pode combinar desfoque, redimensionamento, ruído, compressão JPEG, uma possível segunda etapa de degradação e filtragem final. A *seed* da degradação é derivada da entrada e da partição. Dessa forma, todos os métodos recebem exatamente as mesmas consultas degradadas e as mesmas sementes.

Para identificação fechada, uma galeria HR fixa é construída para cada partição e reutilizada sem alteração por todos os métodos. As consultas de identificação são disjuntas das respectivas imagens de galeria. Falhas no preparo da galeria são preservadas e o número esperado e o número válido de referências são reportados separadamente.

3.1. Métodos avaliados

Foram avaliados os seguintes métodos sobre a mesma lista ordenada de entradas:

- ampliação por vizinho mais próximo;
- interpolação bilinear;

- interpolação bicúbica;
- RRDBNet-base, treinada com perda Charbonnier e termo de fase Fourier;
- RRDBNet com ajuste estrutural;
- RRDBNet com ajuste perceptual VGG, adotada como *baseline* aprendido;
- RRDBNet com ajuste perceptual VGG e perda de identidade baseada em FaceNet.

A variante com identidade é comparada primariamente com a RRDBNet+VGG. O FaceNet participa somente do cálculo da perda de identidade durante o ajuste fino da SR. Ele não é o reconhecedor usado para produzir os resultados de verificação e identificação.

3.2. Pipeline facial

Para cada método, a ordem de processamento é: consulta LR, ampliação ou SR, detecção facial, recorte quadrado, estimação de cinco *landmarks* e alinhamento com MTCNN, extração do *embedding* pelo reconhecedor do InsightFace e, por fim, identificação ou verificação. O reconhecedor efetivamente carregado é o arquivo `w600k_r50.onnx`; ONNX é o formato de serialização e o ONNX Runtime é o mecanismo de execução.

O pipeline registra separadamente ausência de detecção, múltiplas detecções, falha de alinhamento, falha de *embedding* e sucesso. Todas as entradas permanecem nos CSVs. Nas métricas operacionais, uma falha é tratada como rejeição, mantendo o denominador comum; ROC, AUC e EER são calculados somente quando existem escores genuínos e impostores válidos, com o respectivo denominador informado.

3.3. Seleção do limiar e análise estatística

Para cada método estimável, o limiar de similaridade é selecionado pelo critério de EER usando apenas os pares de validação. O limiar e os *hashes* dos escores de validação e da configuração são congelados antes do teste. No partição DevTest da View 1, o limiar é aplicado sem nova otimização e são avaliados pontos operacionais definidos por FAR na validação.

São calculados PSNR, SSIM, MS-SSIM, LPIPS, taxa de detecção, taxa de alinhamento, similaridade de cosseno, distância L_2, Top-1, Top-5, ROC, AUC, EER, FPR, FNR, acurácia, acurácia balanceada, TAR nas FARs predefinidas e por método, imagem e face. Os intervalos de confiança são obtidos por *bootstrap* agrupado por identidade. As diferenças binárias são avaliadas pelo teste exato de McNemar e as diferenças contínuas pelo teste pareado de Wilcoxon, com correção de Holm dentro de cada família de comparações. Além dos valores de p, são reportados a diferença absoluta, o intervalo de confiança e o tamanho do efeito.

3.4. Fluxo de Execução

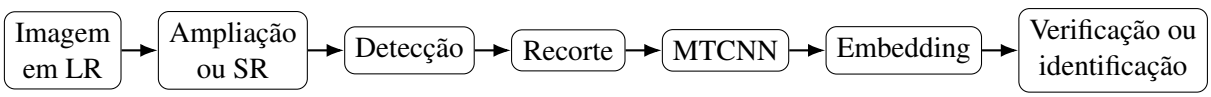


Figura 1. Fluxo comum aplicado a todas as consultas.

O fluxo da Figura 1 é mantido idêntico entre os métodos. Assim, as diferenças observadas podem ser associadas ao método de ampliação ou Super-Resolução, enquanto detecção, seleção da face principal, alinhamento e reconhecimento permanecem fixos.

4. Resultados e discussões

Para verificação, todos os métodos foram avaliados sobre os mesmos pares de validação e teste. A Tabela 1 apresenta as métricas de curva sobre os escores válidos e as métricas operacionais no limiar selecionado exclusivamente na validação. No cálculo operacional, falhas de detecção, alinhamento ou extração do *embedding* permanecem como rejeições.

Tabela 1. Verificação facial no partição DevTest da View 1.

Método	AUC	EER val.	FPR teste	FNR teste	Acc. bal.
bicubic	0.9842	0.0456	0.0400	0.1020	0.9290
bilinear	0.9870	0.0433	0.0400	0.1040	0.9280
nearest	–	–	–	–	–
rrdbnet$_{base}$	0.9838	0.0507	0.0480	0.0640	0.9440
rrdbnet$_{structure}$	0.9820	0.0497	0.0420	0.0660	0.9460
rrdbnet$_{vgg}$	0.9803	0.0489	0.0500	0.0760	0.9370
rrdbnet$_{vgg-identity}$	0.9794	0.0487	0.0440	0.0840	0.9360

O vizinho mais próximo permaneceu na avaliação, mas não produziu escores genuínos e impostores suficientes na validação para estimar ROC, AUC, EER ou limiar. Esses campos são apresentados como não estimáveis, sem substituição por zero e sem seleção de limiar no teste. Suas falhas e métricas visuais continuam reportadas.

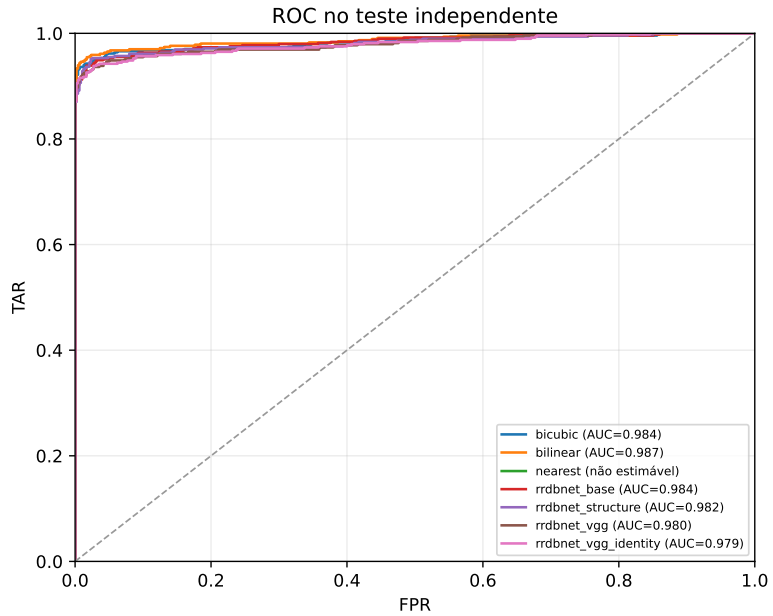


Figura 2. Curvas ROC no partição DevTest da View 1.

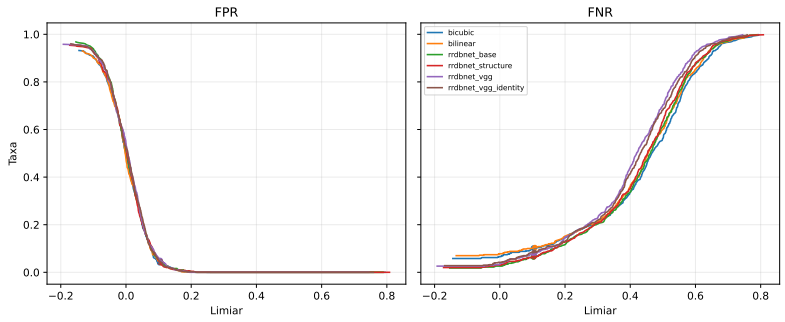


Figura 3. FPR e FNR em função do limiar no teste.

A identificação usa a mesma galeria HR para todos os métodos e mantém as consultas no denominador. A Tabela 2 reporta consultas, *embeddings*, Top-1 e Top-5.

Tabela 2. Identificação no partição DevTest da View 1, com galeria HR comum a todos.

Método	Consultas	Emb. válidos	Top-1	Top-5
bicubic	586	558	0.8652	0.8942
bilinear	586	550	0.8601	0.8908
nearest	586	0	0.0000	0.0000
rrdbnet$_{base}$	586	578	0.8840	0.9215
rrdbnet$_{structure}$	586	576	0.8823	0.9164
rrdbnet$_{vgg}$	586	572	0.8669	0.9027
rrdbnet$_{vgg-identity}$	586	570	0.8720	0.8993

A Tabela 3 reúne qualidade de reconstrução, de detecção, alinhamento e *embedding*. Os valores são agregados a partir das linhas individuais do teste e acompanhados dos respectivos intervalos de confiança.

Tabela 3. Qualidade visual, processamento facial e latência no partição DevTest da View 1.

(a) Qualidade de reconstrução: estimativa [IC95%]

Método	N	PSNR	SSIM	MS-SSIM	LPIPS
Vizinho mais próximo	954	24.63 [24.47; 24.77]	0.7062 [0.7003; 0.7120]	0.8895 [0.8850; 0.8938]	0.5105 [0.5042; 0.5171]
Bilinear	954	25.37 [25.21; 25.54]	0.7529 [0.7470; 0.7589]	0.8948 [0.8905; 0.8990]	0.3849 [0.3769; 0.3928]
Bicúbica	954	26.15 [25.96; 26.34]	0.7670 [0.7608; 0.7734]	0.9046 [0.9002; 0.9090]	0.3689 [0.3605; 0.3771]
RRDBNet-base	954	27.02 [26.82; 27.21]	0.7859 [0.7797; 0.7922]	0.9166 [0.9122; 0.9208]	0.2902 [0.2818; 0.2986]
RRDBNet+estrutura	954	26.99 [26.79; 27.19]	0.7849 [0.7787; 0.7910]	0.9158 [0.9114; 0.9200]	0.2800 [0.2719; 0.2885]
RRDBNet+VGG	954	27.26 [27.05; 27.45]	0.7893 [0.7831; 0.7954]	0.9169 [0.9127; 0.9210]	0.2423 [0.2343; 0.2504]
RRDBNet+VGG+identidade	954	27.37 [27.16; 27.57]	0.7923 [0.7860; 0.7981]	0.9188 [0.9144; 0.9229]	0.2302 [0.2223; 0.2384]

(b) Pipeline facial: estimativa [IC95%]

Método	Detecção	Alinhamento	Embedding	Cosseno HR	L_2 HR
Vizinho mais próximo	0.00% [0.00; 0.00]	0.00% [0.00; 0.00]	0.00% [0.00; 0.00]	–	–
Bilinear	94.13% [92.58; 95.60]	93.40% [91.73; 94.97]	93.40% [91.73; 94.97]	0.6707 [0.6585; 0.6827]	0.7846 [0.7709; 0.7985]
Bicúbica	94.97% [93.51; 96.31]	94.34% [92.80; 95.79]	94.34% [92.80; 95.79]	0.6846 [0.6717; 0.6976]	0.7620 [0.7469; 0.7769]
RRDBNet-base	98.32% [97.49; 99.07]	97.90% [96.95; 98.74]	97.90% [96.95; 98.74]	0.6612 [0.6482; 0.6744]	0.7936 [0.7787; 0.8082]
RRDBNet+estrutura	98.43% [97.61; 99.16]	97.27% [96.19; 98.25]	97.27% [96.19; 98.25]	0.6568 [0.6440; 0.6692]	0.8006 [0.7866; 0.8147]
RRDBNet+VGG	97.80% [96.85; 98.67]	96.86% [95.73; 97.91]	96.86% [95.73; 97.91]	0.6174 [0.6048; 0.6297]	0.8525 [0.8393; 0.8657]
RRDBNet+VGG+identidade	97.80% [96.82; 98.67]	96.75% [95.60; 97.82]	96.75% [95.60; 97.82]	0.6343 [0.6213; 0.6471]	0.8299 [0.8161; 0.8440]

A relação entre qualidade visual e desempenho biométrico na Figura 4 relaciona, por entrada válida e método, métricas visuais e similaridade de cosseno com a referência

HR. As falhas não aparecem como pontos artificiais no gráfico, mas suas quantidades permanecem contabilizadas nas taxas operacionais e nas tabelas.

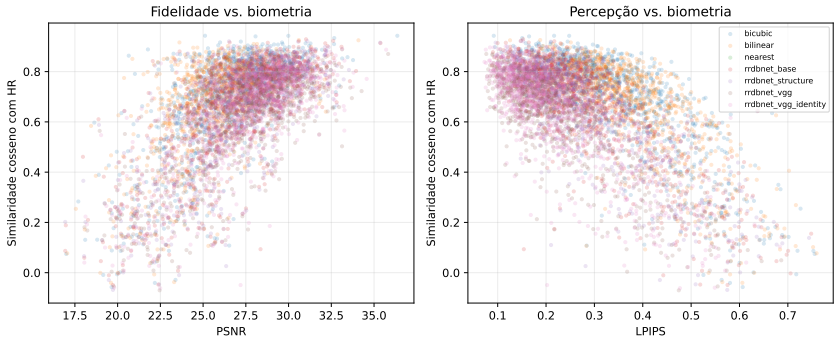


Figura 4. Relação entre fidelidade visual e biométrica para *embedding* válido.

Na comparação primária com a RRDBNet+VGG, a variante com identidade apresentou melhora nas métricas de reconstrução e na proximidade ao *embedding* HR. Entretanto, as diferenças na verificação, Top-1 e Top-5 não foram estatisticamente demonstráveis após a correção de Holm, e os intervalos de confiança das diferenças incluíram zero. A Tabela 4 mostra diferenças, intervalos de confiança, efeito e valores de p corrigidos.

Tabela 4. Comparação entre RRDBNet+VGG+identidade e o RRDBNet+VGG.

Métrica	Teste	N	Diferença [IC95%]	Tamanho de efeito	p_{Holm}
Acerto na verificação	McNemar exato	1000	-0.10 pp [-1.39; 1.28]	OR pareada=0.900	1.0000
PSNR	Wilcoxon pareado	954	0.112 [0.104; 0.121]	r_{rb}=0.804	8.72×10^{-102}
SSIM	Wilcoxon pareado	954	0.0030 [0.0028; 0.0032]	r_{rb}=0.871	6.31×10^{-119}
MS-SSIM	Wilcoxon pareado	954	0.0019 [0.0018; 0.0020]	r_{rb}=0.931	1.15×10^{-135}
LPIPS	Wilcoxon pareado	954	-0.0121 [-0.0125; -0.0117]	r_{rb}=-0.986	6.00×10^{-152}
Cosseno com HR	Wilcoxon pareado	917	0.0164 [0.0138; 0.0188]	r_{rb}=0.633	2.56×10^{-61}
Distância L_2 à HR	Wilcoxon pareado	917	-0.0221 [-0.0247; -0.0193]	r_{rb}=-0.659	2.71×10^{-66}
Acerto Top-1	McNemar exato	586	0.51 pp [-0.74; 1.85]	OR pareada=1.600	1.0000
Acerto Top-5	McNemar exato	586	-0.34 pp [-1.41; 0.68]	OR pareada=0.667	1.0000

Esses resultados mostram uma desconexão mensurável entre os níveis avaliados: maior fidelidade visual e maior proximidade ao *embedding* da referência não implicaram melhora estatisticamente demonstrável nas decisões finais de verificação ou identificação. Essa conclusão é restrita ao protocolo, às degradações, aos reconhecedores e aos métodos avaliados; ela não constitui uma afirmação universal sobre Super-Resolução e biometria.

5. Conclusões e trabalhos futuros

Os experimentos mostram que a avaliação de Super-Resolução para reconhecimento facial deve considerar a cadeia completa de processamento, e não somente métricas de reconstrução. No protocolo independente adotado, a variante com perda de identidade melhorou a fidelidade visual e a proximidade ao *embedding* referência em relação ao *baseline* RRDBNet+VGG. Entretanto, essa melhoria não produziu ganho estatisticamente demonstrável na verificação, na identificação Top-1 ou Top-5.

Portanto, os resultados sustentam uma conclusão limitada: para os métodos avaliados, melhoria visual e maior proximidade ao *embedding* HR pode não implicar na a

melhor decisão biométrica. As falhas de detecção e alinhamento afetam o desempenho fim-a-fim e, por isso, foram preservadas no denominador em vez de descartadas. Desta forma, por si só, a aplicabilidade do sistema ao controle de presença em sala de aula.

Como trabalhos futuros, podem ser avaliados métodos especializados de restauração facial, caso sua viabilidade técnica e seus pesos sejam verificados, e uma perda de identidade calculada no mesmo espaço do reconhecedor final.

Referências

Adjabi, I., Ouahabi, A., Benzaoui, A., and Taleb-Ahmed, A. (2020). Past, present, and future of face recognition: A review. *Electronics*, 9(8).

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press, Cambridge, Massachusetts.

He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition.

Huang, G. B., Ramesh, M., Berg, T., and Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical report, University of Massachusetts, Amherst.

InsightFace (2026). Insightface: 2d and 3d face analysis project. `https://github.com/deepinsight/insightface`. Accessed: 14 Aug. 2026.

Johnson, J., Alahi, A., and Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711.

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

ONNX Community (2019). Open neural network exchange (onnx). `https://github.com/onnx/onnx`. Accessed: 2026-08-06.

Schroff, F., Kalenichenko, D., and Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 815–823.

Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. (2014). Deepface: Closing the gap to human-level performance in face verification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1701–1708. IEEE.

Turk, M. and Pentland, A. (1991). Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86.

Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., and Loy, C. C. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks. In *European Conference on Computer Vision Workshops*.

Zhang, K., Zhang, Z., Li, Z., and Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Processing Letters*, 23(10):1499–1503.