

Melhoria de Qualidade de Imagens usando CNNs

Alan P. I. R. Gonçalves, Gustavo de O. F. Duarte, Rafael T. C. dos Santos, Pedro de C. C. Pinto, Gustavo M. S. Nunes, Fernanda D. V. R. Oliveira, José Gabriel R. C. Gomes

Resumo— Neste trabalho, analisamos o desempenho de redes neurais convolucionais para a melhoria da qualidade de imagens degradadas por ruído e compressão. Treinamos as redes autoencoder, autoencoder residual e U-Net visando otimizar as métricas de qualidade SSIM e 3-SSIM. Melhorias entre 5 e 9 pontos percentuais (pp) na SSIM em relação a uma filtragem gaussiana indicam que os resultados das redes são mais similares à imagem original. Observamos que redes convolucionais maiores, com mais parâmetros, tendem a gerar imagens de melhor qualidade.

Palavras-Chave— Imagens, Ruído, Redes Neurais, Autoencoder, U-Net, SSIM, PSNR, PSNR-B, 3-SSIM.

Abstract— In this work, we analyze convolutional neural network (CNN) performances in image denoising. We trained autoencoder, residual autoencoder, and U-Net models to optimize 3-SSIM and SSIM quality metrics. SSIM improvements between 5 and 9 percentage points in comparison to the gaussian filter indicate that results obtained from the CNNs have higher similarity to the pristine images. We observe that deeper CNNs, with more parameters, tend to generate better quality images.

Keywords— Images, Noise, Denoising, CNN, Autoencoder, U-Net, SSIM, PSNR, PSNR-B, 3-SSIM.

I. INTRODUÇÃO

A motivação para este estudo é um problema de degradação das imagens capturadas pelo sensor CMOS de [1], causada por compressão analógica e ruído. O artigo [2] descreve técnicas de aprendizado profundo na melhoria da qualidade de imagens. Neste artigo realizamos uma análise de redes convolucionais na melhoria da qualidade de imagens comprimidas digitalmente e com ruído acrescentado de forma artificial.

Em nosso estudo, a entrada de uma rede autoencoder [3] é uma imagem em tons de cinza, comprimida e com ruído. O autoencoder deve extrair as características da imagem e, com elas, gerar uma nova imagem, de mesma dimensão, e com qualidade melhorada. De forma semelhante, o autoencoder residual é uma versão modificada, através da adição de conexões residuais entre as camadas, que também é utilizada para remoção de ruído. Na rede U-Net [4], as dimensões da saída também são iguais às da entrada. Contudo, a U-Net reduz o efeito de perda de características a cada convolução através de conexões diretas entre camadas da rede.

Utilizando as métricas *structural similarity index measure* (SSIM) [6], *three-component SSIM* (3-SSIM) [7], e *peak signal-to-noise ratio* (PSNR) *including blocking effect* (PSNR-B) [8], quantificamos a melhora da qualidade de uma imagem e comparamos os resultados das redes com os resultados de uma filtragem gaussiana [5]. A 3-SSIM permite ponderar a

SSIM em diferentes regiões da imagem, avaliando com resultados mais altos as imagens que possuem mais semelhança em regiões de borda. A PSNR-B é uma modificação da PSNR que inclui um fator de efeito de blocagem, que é um tipo de distorção causado pela compressão por blocos.

II. CONFIGURAÇÕES E RESULTADOS DOS TESTES

Realizamos dois testes. No teste I, treinamos autoencoders e U-Nets por 50 épocas para maximizar a SSIM das imagens da base CIFAR-10. Comprimos as imagens com JPEG, e introduzimos ruído branco aditivo gaussiano (*additive white gaussian noise* - AWGN) a cada uma delas. Para cada rede, permutamos os valores de compressão JPEG (30:1 e nenhuma) e de desvio padrão de ruído (0.1 e 0.3), gerando combinações diferentes de entradas. Para cada combinação, consideramos diferentes configurações de hiper-parâmetros, visando obter os melhores desempenhos de treinamento.

Com o objetivo de avaliar os modelos em um cenário mais próximo do problema com o circuito CMOS, no teste II, transformamos as imagens da base CIFAR-10 para a escala de cinza. Em seguida, aumentamos sua resolução para 64×64 pixels. Por fim, comprimimos as imagens com o algoritmo equivalente ao realizado pelo circuito CMOS mostrado no artigo [1]. Posteriormente, treinamos 31 modelos para maximizar a métrica 3-SSIM. Os modelos incluem U-Nets, autoencoders, autoencoders residuais, e redes formadas apenas por camadas convolucionais e residuais (sem *pooling*). Avaliamos os desempenhos dos modelos pelas métricas 3-SSIM, SSIM e PSNR-B. O treino de cada modelo dura até que ele não obtenha um progresso maior do que 0.2% na métrica 3-SSIM, em uma janela de sete épocas.

A. Resultados do teste I

A Tabela I mostra a média dos valores de SSIM entre imagens de saída de cada modelo e as imagens originais (coluna SSIM 2), para diferentes níveis de compressão e ruído. Para permitir a visualização da melhora de qualidade, a coluna SSIM 1 mostra a média dos valores de SSIM entre imagens de entrada e imagens originais. A U-Net e o autoencoder apresentam resultados de SSIM superiores aos do filtro gaussiano em pelo menos 5 pontos percentuais (pp). Além disso, a U-Net desempenha melhor do que o autoencoder na base utilizada. Por exemplo, nas 3 primeiras linhas da tabela, a diferença de desempenho entre a U-Net e o autoencoder é de quase 2 pp.

A Figura 1 ilustra duas imagens do primeiro conjunto de teste em quatro situações: imagem original, imagem comprimida e com ruído, imagem gerada pela U-Net após o treino, e imagem gerada pelo filtro gaussiano. Seus respectivos valores de SSIM com relação à imagem original são apresentados na

Todos os autores são filiados à Universidade Federal do Rio de Janeiro, Departamento de Eng. Eletrônica e Programa de Engenharia Elétrica. E-mail para contato: fernanda.dvro@poli.ufrj.br. Este trabalho foi financiado pelas agências brasileiras de incentivo à pesquisa FAPERJ, CAPES (Código de Financiamento 001) e CNPq (projetos 309089/2022-0 e 142264/2019-9).

TABELA I
RESULTADOS OBTIDOS DO TESTE I.

Arquitetura	Compr. (JPEG)	Ruído (AWGN)	SSIM 1	SSIM 2
U-Net	30	0.3	0.2457	0.6556
Autoencoder	30	0.3	0.2455	0.6373
Filtro Gaussiano	30	0.3	0.2453	0.5687
U-Net	30	0.1	0.6056	0.8549
Autoencoder	30	0.1	0.6052	0.8417
Filtro Gaussiano	30	0.1	0.6056	0.7940
U-Net	-	0.3	0.2617	0.6622
Autoencoder	-	0.3	0.2615	0.6374
Filtro Gaussiano	-	0.3	0.2614	0.5781
U-Net	-	0.1	0.6492	0.8763
Autoencoder	-	0.1	0.6492	0.8611
Filtro Gaussiano	-	0.1	0.6491	0.8122

figura. A imagem de saída da rede apresenta maior nitidez do que a imagem comprimida e com ruído. Isto ilustra que a rede está, de fato, conseguindo melhorar a qualidade da imagem degradada. Em comparação com a imagem gerada pelo filtro gaussiano, a imagem de saída da rede também é mais nítida e tem qualidade mais próxima à da imagem original.

B. Resultados do teste II

A Figura 2 mostra os melhores resultados obtidos considerando as métricas SSIM, 3-SSIM e PSNR-B variando o número de parâmetros de cada modelo. Conforme o número de parâmetros dos modelos aumenta, as imagens geradas por eles apresentam, em média, maior qualidade. Já quando a ordem do número de parâmetros do modelo ultrapassa 10^5 , notamos uma estagnação no valor das métricas. Possivelmente, isto é causado por uma quantidade insuficiente de dados para treinar modelos deste tamanho, resultando em perda de generalização.

Na Figura 3, há uma comparação entre imagens originais (sem ruído), imagens comprimidas, imagens de saída de um autoencoder residual (com 166.711 parâmetros), e imagens de saída do filtro gaussiano. Os valores de 3-SSIM, SSIM e PSNR-B, calculados em relação à imagem original, são exibidos na figura. Comparado ao filtro gaussiano, o autoencoder residual obtém resultados superiores em todas as métricas de qualidade consideradas: ao menos 11pp para 3-SSIM, 12pp para SSIM, e 7dB para PSNR-B. Essa melhora é notada, também, visualmente, com imagens do autoencoder residual sendo mais nítidas e tendo qualidade mais próxima à imagem original. Para todas as imagens da base de teste, o autoencoder residual considerado na Figura 3 obteve uma média de 0.882 na métrica 3-SSIM.

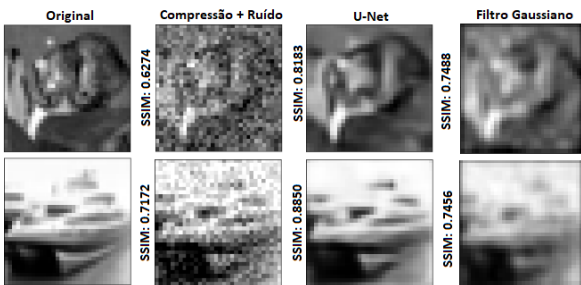


Fig. 1. Imagens originais, com compressão e ruído, saídas da U-Net, e saídas do filtro gaussiano extraídas do teste I.

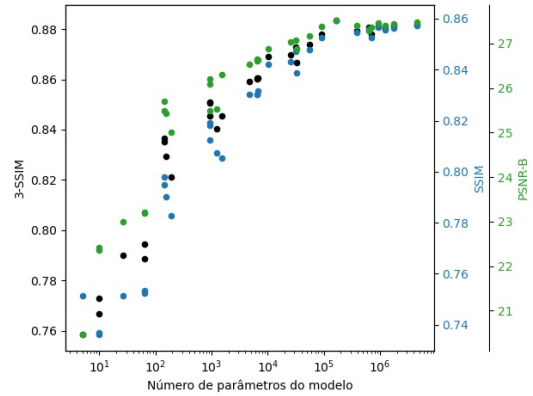


Fig. 2. Melhores resultados de média, no conjunto de teste II, nas métricas PSNR-B (verde), SSIM (azul) e 3-SSIM (preto) de cada modelo (eixo y), em relação ao número de parâmetros da rede (eixo x, em escala logarítmica).

III. CONCLUSÕES

Neste trabalho, mostramos que modelos baseados em redes convolucionais são capazes de melhorar significativamente a qualidade das imagens, obtendo resultados melhores que filtros gaussianos. Quando testamos em um problema semelhante ao real, conseguimos resultados satisfatórios. Também evidenciamos uma relação entre o número de parâmetros de diversas arquiteturas diferentes e os valores das métricas de qualidade.

REFERÊNCIAS

- [1] F. D. V. R. Oliveira et al. "Focal-Plane Compression Imager with Increased Quantization Bit Rate and DPCM Error Modeling", Journal of Integrated Circuits and Systems 2017; v.12 / n.2:71-81
- [2] C. Tian et al., "Deep Learning on Image Denoising: An overview". arXiv, 3 de agosto de 2020. Acessado: 10 de outubro de 2022. arXiv:1912.13171.
- [3] D. Bank et al., "Autoencoders", arXiv:2003.05991, 2021.
- [4] O. Ronneberger et al., "U-Net: Convolutional Networks for Biomedical Image Segmentation", arXiv:1505.04597, 2015.
- [5] G. Deng and L. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection", Proc. IEEE NSSMIC, pp. 1615-1619, 1994.
- [6] J. Nilsson, T. Akenine-Möller, "Understanding SSIM", arXiv:2006.13846, 2020
- [7] A. C. Bovik, "Content-weighted video quality assessment using a three-component image model", J. Electron. Imaging, vol. 19, nº 1, p. 011003, jan. 2010.
- [8] Changhoon Yim and A. C. Bovik, "Quality Assessment of Deblocked Images", IEEE Trans. Image Proc., vol. 20, nº 1, p. 88-98, jan. 2011.

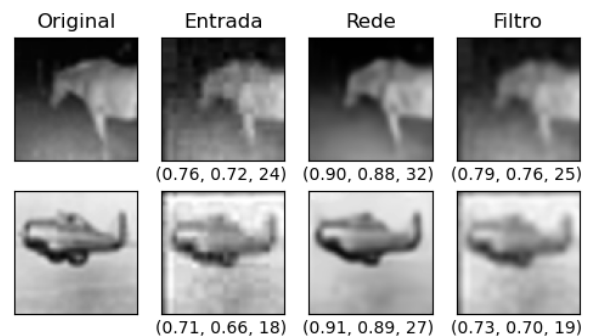


Fig. 3. Comparação entre imagens originais, imagens de contaminadas com ruído, saída da rede autoencoder residual, saída do filtro gaussiano, do teste II: (3-SSIM, SSIM, PSNR-B) da imagem em questão em relação à original.