Universidade Federal do ABC
Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas
Bacharelado em Engenharia de Informação

Aprimoramento de imagens com Redes Neurais Artificiais no domínio da Transformada Discreta de Cosseno

Adinan Alves de Brito Filho

Santo André - SP
Dezembro de 2025

Adinan Alves de Brito Filho

Aprimoramento de imagens com Redes Neurais Artificiais no domínio da Transformada Discreta de Cosseno

Trabalho de Graduação apresentado ao curso de Engenharia de Informação da Universidade Federal do ABC, como parte dos requisitos para a obtenção do Título de Bacharel em Engenharia de Informação.

Universidade Federal do ABC – UFABC

Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas

Orientador: Prof. Dr. Kenji Nose Filho

Santo André - SP

Dezembro de 2025

Adinan Alves de Brito Filho

Aprimoramento de imagens com Redes Neurais Artificiais no domínio da Transformada Discreta de Cosseno

Trabalho de Graduação apresentado ao curso de Engenharia de Informação da Universidade Federal do ABC, como parte dos requisitos para a obtenção do Título de Bacharel em Engenharia de Informação.

Santo André, 26 de novembro de 2025.

Prof. Dr. Kenji Nose Filho
Orientador

Prof. Dr. Ricardo Suyama
Avaliador 1

Prof. Dr. Murilo Bellezoni Loiola
Avaliador 2

Santo André - SP
Dezembro de 2025

AGRADECIMENTOS

Agradeço, primeiramente, ao Senhor Jesus Cristo, autor e consumador da minha fé, por me conceder força, sabedoria e perseverança ao longo de toda a minha trajetória, e por nunca desistir de mim.

À minha família, pelo amor, pelo encorajamento constante, pelo apoio irrestrito e pelo genuíno interesse demonstrado em cada etapa deste trabalho e da minha formação acadêmica.

Ao meu orientador, Prof. Dr. Kenji Nose Filho, por acreditar em mim e investir em meu desenvolvimento desde o primeiro dia, no início de 2019; por tanto ensinamento, pelos conselhos valiosos e pela empatia demonstrada, que me marcaram e marcam a todos ao seu redor.

Ao meu Pr. Agnaldo Araujo, pelo amor e cuidado espiritual dedicados às suas ovelhas e por nos ensinar, na prática, o verdadeiro exemplo de ser cristão.

Aos meus amigos, que tornaram esta jornada ainda mais leve e prazerosa. Como bem expressa Asaph Borba: *“Quem anda sozinho pode ir mais rápido, mas nem sempre vai mais longe.”*

Aos meus colegas na IBM, especialmente Marcelo Grave e Cassia Sampaio, pelo acompanhamento, apoio e suporte ao longo deste período. Trabalhar com pesquisa aplicada representou a realização de um sonho e foi fundamental para o amadurecimento deste trabalho.

Por fim, agradeço a Tito Spadini e a toda a comunidade acadêmica da Engenharia de Informação — professores e estudantes — que constroem diariamente um curso singular, reconhecido não apenas pela excelência técnica, mas também pela formação humana que promove.

“Pois todas as coisas vêm d’Ele,
existem por meio d’Ele e são para Ele.
A Ele seja toda a glória para sempre! Amém.“
Romanos 11:36

RESUMO

Este trabalho apresenta o desenvolvimento de um método para aprimoramento de imagens degradadas por compressão com perdas, como no formato JPEG, baseado em Redes Neurais Artificiais e na Transformada Discreta de Cosseno (DCT), intitulado `ContextResidualMLP`. A abordagem proposta atua diretamente no domínio da DCT, aprendendo, a partir da comparação entre coeficientes de imagens comprimidas e de alta qualidade, padrões estatísticos de degradação, de modo a atenuar artefatos introduzidos pela compressão. Experimentos conduzidos sob diferentes níveis de qualidade JPEG indicaram melhorias consistentes em métricas objetivas, como aumentos de PSNR e MS-SSIM, bem como reduções no NIQE. Tais resultados evidenciam a contribuição da solução proposta para a mitigação dos efeitos da compressão com perdas, favorecendo a preservação da qualidade visual em imagens digitais.

Palavras-chave: Aprimoramento de imagens. Transformada Discreta de Cosseno (DCT). Redes Neurais Artificiais (RNA). *Multilayer Perceptron* (MLP). Inteligência artificial aplicada. Processamento de sinais multimídia.

ABSTRACT

This work presents the development of a method for enhancing images degraded by lossy compression, such as in the JPEG format, based on Artificial Neural Networks and the Discrete Cosine Transform (DCT), referred to as `ContextResidualMLP`. The proposed approach operates directly in the DCT domain, learning statistical degradation patterns from the comparison between compressed and high-quality image coefficients in order to attenuate compression-induced artifacts. Experiments conducted under different JPEG quality levels indicated consistent improvements in objective metrics, including increases in PSNR and MS-SSIM, as well as reductions in NIQE. These results demonstrate the contribution of the proposed solution to mitigating the effects of lossy compression, thereby promoting the preservation of visual quality in digital images.

Keywords: Image enhancement. Discrete Cosine Transform (DCT). Artificial Neural Networks (ANN). Multilayer Perceptron (MLP). Artificial intelligence applications. Multimedia signal processing.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AAC	Advanced Audio Coding
BRISQUE	Blind/Referenceless Image Spatial Quality Evaluator
CCITT	International Telegraph and Telephone Consultative Committee
CSV	Comma-separated values
DCT	Discrete Cosine Transform
GELU	Gaussian Error Linear Unit
IEC	International Electrotechnical Commission
ISO	International Organization for Standardization
ITU	International Telecommunications Union
HDTV	High-definition television
HEIF	High Efficiency Image File Format
JPEG	Joint Photographic Experts Group
MAE	Mean Absolute Error
ML	Machine Learning
MLP	Multilayer Perceptron
MPEG	Moving Picture Experts Group
MS-SSIM	Multiscale Structural Similarity Index Measure
MSE	Mean Squared Error
NIQE	Natural Image Quality Evaluator
PNG	Portable Network Graphics
PSNR	Peak Signal-to-Noise Ratio
RGB	Red, Green, Blue
RLE	Run-Length Encoding
SDTV	Standard-definition television
SSIM	Structural Similarity Index Measure
YCbCr	Luma, Blue-difference Chroma, Red-difference Chroma

SUMÁRIO

1 INTRODUÇÃO

Os sinais multimídia constituem uma classe de sinais que abrange, de modo geral, conteúdos de áudio, fala, imagem e vídeo. As comunicações multimídia, por sua vez, englobam processos de geração, armazenamento e transmissão dessa classe de sinais (HWANG, 2009; RUSS, 2011).

Desde meados do século XX, com o avanço tecnológico e o surgimento dos meios digitais, as comunicações multimídia ganharam destaque no mundo contemporâneo. A cada ano, cresce o volume de conteúdo produzido, armazenado e distribuído pela internet e radiodifusão. Entretanto, grande parte do acervo visual das últimas décadas foi preservada em baixa resolução e sob altos graus de compressão, devido às limitações técnicas e de custo vigentes em cada época.

Com a evolução das tecnologias de captura, produção e transmissão de sinais multimídia, muitos padrões antes predominantes tornaram-se defasados, frequentemente incompatíveis com as expectativas atuais de consumo e com as práticas industriais. Pesquisas recentes indicam que a qualidade técnica (resolução, taxa de bits e condições de transmissão) influencia tanto a compreensão do conteúdo (DONTRE, 2023) quanto a Qualidade de Experiência (QoE) do usuário, afetando diretamente o engajamento e os resultados dos negócios (TECHNOLOGIES, 2020).

Diante da relevância da qualidade técnica para a experiência do consumo, da defasagem dos padrões legados e da importância histórica e cultural do acervo produzido, torna-se necessária a atualização e o aprimoramento de conteúdos mantidos em baixa resolução — sobretudo aqueles sujeitos à compressão com perdas — de modo a mitigar a degradação e compatibilizá-los a padrões contemporâneos, contribuindo, assim, para a disseminação e a preservação do patrimônio visual produzido ao longo das últimas décadas (AYDIN; CİNBİŞ, 2020).

Para contribuir na resolução desses desafios, diversas técnicas de aprimoramento de imagens têm sido propostas, incluindo métodos baseados em interpolação, filtragem e aprendizado de máquina (DONG et al., 2014; DONG et al., 2015; ZAMIR et al., 2022). Dentre essas abordagens, redes neurais profundas têm se destacado pela capacidade de aprender padrões complexos e generalizar para novos dados (GOODFELLOW; BENGIO; COURVILLE, 2016; LECUN; BENGIO; HINTON, 2015).

Neste trabalho, investigamos o aprimoramento de imagens no domínio espectral a partir da Transformada Discreta de Cosseno (DCT). A DCT é um dos pilares da compressão de imagens, sustentando padrões amplamente utilizados, como o JPEG (*Joint Photographic Experts Group*) (WALLACE, 1992), e influenciando fortemente padrões de vídeo, como os do *Moving Picture Experts Group* (MPEG) (GALL, 1991),

que viabilizaram o armazenamento e a distribuição em larga escala desde a década de 1990 e permanecem em uso até hoje.

A opção pela DCT como domínio para o aprimoramento decorre de sua relevância prática: ao representar a imagem no espectro, grande parte da informação visual concentra-se em poucos coeficientes de baixa frequência (AHMED; NATARAJAN; RAO, 1974). A compressão reduz a magnitude dos coeficientes de alta frequência (tipicamente menos relevantes para a percepção humana), diminuindo assim a quantidade de dados necessária. Contudo, tal compressão com perdas introduz artefatos visuais indesejados — como *blocking* e *ringing* — que degradam a qualidade perceptual.

Com o objetivo de aprimorar imagens que tenham sido degradadas devido à compressão com perdas do JPEG, o presente trabalho propõe uma arquitetura de rede neural denominada `ContextResidualMLP`. O treinamento é conduzido com pares de dados formados por coeficientes DCT obtidos de imagens degradadas por compressão JPEG (baseada em DCT) e os coeficientes correspondentes de suas imagens de referência em alta qualidade. A rede, do tipo *Multi-Layer Perceptron* (MLP), é treinada para estimar o resíduo de correção a ser somado aos coeficientes degradados — minimizando o erro em relação ao alvo. Dessa forma, busca-se atenuar os efeitos introduzidos, principalmente, pela quantização no domínio transformado, realizando a restauração diretamente no domínio em que a degradação é gerada. Ao operar sobre representações no domínio da frequência e empregar aprendizado residual, o método visa reduzir artefatos de compressão e limitar alterações indevidas no conteúdo, promovendo maior fidelidade à imagem de referência.

Além disso, como parte das contribuições deste trabalho, desenvolveu-se uma aplicação web interativa para o aprimoramento de imagens por meio dos modelos de redes neurais da arquitetura proposta. A aplicação permite o envio de imagens de teste, realiza o aprimoramento utilizando os modelos desenvolvidos e apresenta uma comparação visual entre a imagem de entrada, degradada, e a imagem aprimorada. Ademais, a aplicação exibe métricas objetivas sem referência (NR, do inglês *No-Reference*), como o NIQE (do inglês Natural Image Quality Evaluator) (MITTAL; SOUNDARARAJAN; BOVIK, 2013) e a entropia (TSAI et al., 2008), que possibilitam a mensuração da qualidade da imagem degradada e da imagem reconstruída, sem a necessidade de comparação com uma versão de alta qualidade. Essa aplicação visa viabilizar a utilização prática do trabalho desenvolvido, além de facilitar a reprodutibilidade e a avaliação independente dos resultados. Detalhes sobre a interface *web* são apresentados no capítulo de Resultados.

Na sequência, apresentam-se os conceitos fundamentais que norteiam este trabalho, bem como os procedimentos experimentais adotados e os resultados obtidos.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são apresentados os principais conceitos e definições que embasam este trabalho, incluindo a definição formal de imagem, a DCT, Redes Neurais Artificiais e as métricas de avaliação utilizadas para mensurar os resultados obtidos. O entendimento desses conceitos é fundamental para a plena compreensão da metodologia proposta e dos resultados alcançados.

2.1 IMAGENS DIGITAIS E REPRESENTAÇÃO

Uma imagem pode ser definida matematicamente como uma função bidimensional, $f(x, y)$, em que x e y são as coordenadas espaciais de um plano, e a amplitude de f em qualquer par de coordenadas (x, y) é chamada de intensidade. Quando x, y e os valores de intensidade de f são quantidades finitas e discretas, temos, então, uma imagem digital, na qual cada par ordenado representa um pixel.

Uma imagem digital é composta por uma matriz bidimensional de píxeis, onde cada pixel representa a menor unidade de informação visual na imagem. A resolução de uma imagem digital é definida pelo número de píxeis na largura e na altura da imagem, geralmente expressa como largura $\times$ altura (por exemplo, 1920×1080 píxeis). Cada pixel possui um valor que determina a intensidade de um canal de cor (no caso de uma imagem RGB) ou o tom de cinza (no caso de uma imagem em escala de cinza) representado por esse pixel. A profundidade de cor, ou *bit depth*, refere-se ao número de bits utilizados para representar a intensidade de cada pixel, influenciando diretamente a gama de cores ou tons que podem ser representados na imagem. Por exemplo, uma profundidade de cor de 8 bits permite representar 256 níveis de intensidade (0-255) para cada canal de cor em imagens RGB, resultando em mais de 16 milhões de cores possíveis.

2.2 O PADRÃO JPEG

O JPEG é um padrão amplamente utilizado para compressão de imagens digitais, especialmente aquelas com muitas cores e detalhes, como fotografias. Foi desenvolvido entre 1985 e 1990 por um comitê conjunto da ISO (*International Organization for Standardization*) e do CCITT (*International Telegraph and Telephone Consultative Committee*), hoje ITU-T, parte da ITU, e publicado como padrão internacional em 1992 (WALLACE, 1992), sendo conhecido formalmente como ISO/IEC 10918-1 e ITU-T T.81, com o objetivo de fornecer um método eficiente de compressão de imagens com perdas, reduzindo o tamanho dos arquivos com um baixo custo computacional, de forma a facilitar o armazenamento e a transmissão de imagens digitais em redes

de computadores e dispositivos com capacidade limitada de armazenamento (ITU-T, 1992).

O JPEG define vários processos de compressão para imagens estáticas. Dentre esses, seu método padrão, o mais utilizado, envolve várias etapas, incluindo a conversão da imagem para o sistema de cores YCbCr, a divisão da imagem em blocos de 8×8 píxeis, a aplicação da DCT em cada bloco, a quantização dos coeficientes DCT, a codificação dos dados quantizados usando técnicas como a Codificação RLE (*Run-Length Encoding*) e a Codificação de Huffman (WALLACE, 1992), *visando mitigar a redundância de bits*. O padrão JPEG é amplamente suportado por diversos softwares e dispositivos, sendo uma escolha popular para armazenamento e compartilhamento de imagens digitais devido à sua eficiência na compressão e qualidade visual.

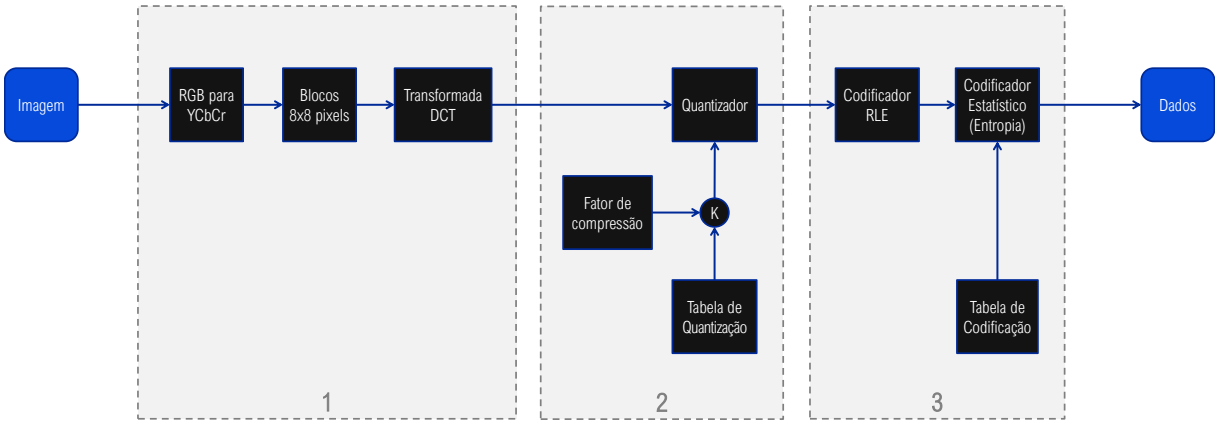
A Figura 1 ilustra um diagrama de blocos de um codificador JPEG padrão. Este diagrama está subdividido em três blocos:

1º: Pré-processamento da imagem;

2º: Compressão com perdas;

3º: Compressão sem perdas e codificação;

Figura 1 – Diagrama de blocos de um codificador JPEG.



Fonte: Adaptação de (STOLFI, 2016).

Parte da metodologia desenvolvida neste trabalho compreende o primeiro e o segundo bloco do diagrama do codificador JPEG, conforme ilustrado na Figura 1. Nas próximas seções, são detalhados os principais conceitos envolvidos nesses blocos.

2.3 SISTEMAS DE CORES

Sistemas de cores são modelos matemáticos que descrevem a maneira como as cores são representadas e percebidas. Eles são essenciais para a reprodução precisa

de cores em diferentes dispositivos, como monitores, impressoras e câmeras digitais. Matematicamente, é um modelo que descreve a representação de cores como uma tupla de números, onde cada elemento pode representar a intensidade de uma cor em específico ou sua tonalidade (GONZALEZ; WOODS, 2018).

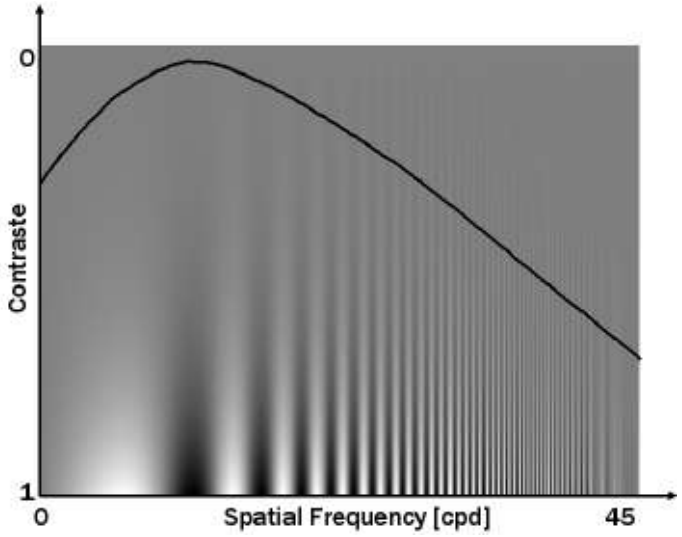
Existem vários sistemas de cores, cada um com suas próprias características e aplicações. Alguns dos sistemas de cores mais comuns incluem:

- RGB (*Red, Green, Blue*): Utilizado principalmente em dispositivos eletrônicos, como monitores e televisores. As cores são representadas pela combinação de diferentes intensidades de vermelho, verde e azul. Caracteriza-se por ser um sistema de cores aditivo, ou seja, as cores são formadas pela adição de luz.
- CMYK (*Cyan, Magenta, Yellow, Key/Black*): Utilizado principalmente na impressão. As cores são representadas pela combinação de ciano, magenta, amarelo e preto. Caracteriza-se por ser um sistema de cores subtrativo, ou seja, as cores são formadas pela subtração de luz.
- YCbCr: Utilizado em compressão de imagens e vídeos, como no padrão JPEG. A componente Y representa a luminância (brilho), enquanto Cb e Cr representam as componentes de crominância (cor).
- HSL (*Hue, Saturation, Lightness*) e HSV (*Hue, Saturation, Value*): Modelos que representam as cores em termos de matiz, saturação e luminosidade ou valor. São frequentemente utilizados em design gráfico e na edição de imagens.

Tipicamente, imagens digitais são representadas no sistema de cores RGB. Seria possível utilizar este sistema de cores para a transformação DCT, porém, vale-se de um aspecto psicofísico do sistema visual humano: o sistema visual humano é mais sensível à intensidade da luz do que a pequenas variações de cromaticidade. Tal fato é evidenciado pela Figura 2, que ilustra a função de sensibilidade ao contraste do sistema visual humano. Ao analisá-la, nota-se que a percepção visual varia conforme a frequência espacial do estímulo: padrões associados a altas frequências espaciais, caracterizados por variações rápidas de intensidade luminosa, requerem maior contraste para serem percebidos, o que justifica a maior tolerância a distorções nessas componentes em técnicas de compressão e processamento no domínio da frequência.

A partir desse aspecto, é possível inferir que se pode realizar uma compressão maior nos canais de crominância (cor) da imagem em comparação ao canal de luminância e, ainda assim, obter um resultado satisfatório, dado que o sistema visual humano tende a perceber mais a perda de informação de luminosidade do que a perda

Figura 2 – Gráfico da Função de Sensibilidade ao Contraste (CSF – *Contrast Sensitivity Function*). O gráfico contém uma representação de padrões em tons de cinza com diferentes frequências espaciais, evidenciando que o contraste mínimo necessário para que o sistema visual humano perceba as variações de intensidade não é constante, sendo maior para frequências espaciais mais altas.



Fonte: Extraído de (HAUTIèRE; TAREL; BRéMOND, 2007).

de informação de cor. Por isso, realiza-se a conversão das imagens para o sistema de cores YCbCr.

A conversão de RGB para YCbCr (HAMILTON, 1992), em uma imagem de 8 bits por canal de cor, ocorre por meio da transformação expressa na Equação 2.1:

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ Cb &= -0,1687R - 0,3313G + 0,5B + 128 \\ Cr &= 0,5R - 0,4187G - 0,0813B + 128 \end{aligned} \tag{2.1}$$

onde R, G e B são os valores das intensidades dos canais vermelho, verde e azul, respectivamente, e Y, Cb e Cr são os valores das intensidades dos canais de luminância e crominância.

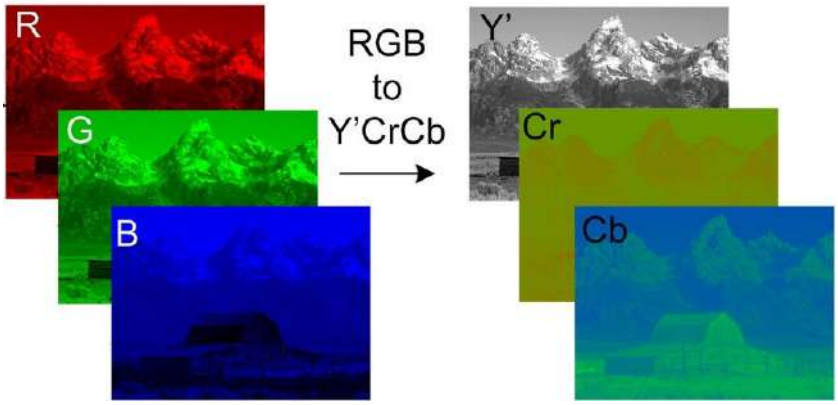
O processo inverso, a conversão do sistema de cores YCbCr para RGB (HAMILTON, 1992), é realizado pela transformação expressa na Equação 2.2.

$$\begin{aligned} R &= Y + 1.402(Cr - 128) \\ G &= Y - 0,344136(Cb - 128) - 0,714136(Cr - 128) \\ B &= Y + 1,772(Cb - 128) \end{aligned} \tag{2.2}$$

Ao final de ambas as conversões, os valores obtidos são arredondados para valores inteiros. A Figura 3 ilustra de maneira visual a transformação do sistema de cores RGB para YCbCr.

Um fator que corrobora a conversão do sistema de cores é que o YCbCr requer menos espaço de memória para sua representação numérica e, consequentemente,

Figura 3 – Figura ilustrativa da conversão do sistema de cores de uma imagem do RGB para o YCbCr.

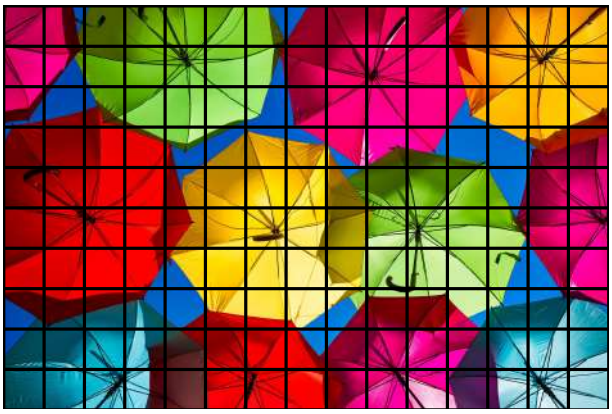


Fonte: Extraído de (ASSI; SHANWAR; ZARKA, 2016).

para seu armazenamento. Isso ocorre porque é possível realizar grandes variações de cores no sistema YCbCr com valores significativamente menores do que os dos canais RGB correspondentes à mesma cor.

Em seguida, cada canal da imagem é particionado em pequenos blocos (subimagens) de 8 por 8 píxeis que não se sobrepõem, e cada subimagem é processada individualmente utilizando uma transformada 2D reversível, que, neste caso, é a DCT. Esta transformação será explicitada a seguir. A Figura 4 ilustra o processo de particionamento.

Figura 4 – Esquema de particionamento da imagem em blocos de 8×8 píxeis. O particionamento é realizado em todos os canais de cor. Figura ilustrativa (fora de escala).



Fonte: De autoria própria.

2.4 A TRANSFORMADA DISCRETA DE COSSENO (DCT)

Transformadas são operadores matemáticos utilizados para facilitar a busca de soluções para problemas não triviais, uma vez que a transformação realizada reduz a complexidade do problema original, tornando-o mais simples de resolver ou mais favorável ao propósito a que se destina (BRACEWELL, 2000). Seu uso é muito

vantajoso, pois é possível resolver o problema transformado e retornar ao seu domínio original aplicando a transformada inversa à solução.

A Transformada Discreta de Cosseno, também conhecida como DCT (*Discrete Cosine Transform*), expressa uma sequência finita de dados em termos de uma soma de funções cosseno em diferentes frequências. A DCT se relaciona com a Transformada de Fourier e é similar à Transformada Discreta de Fourier (DFT - *Discrete Fourier Transform*); no entanto, utiliza apenas o conjunto dos números reais (LIVERA; RIBEIRO, 2014).

A DCT foi proposta por Nasir Ahmed em 1972, durante sua estadia na Universidade Estadual de Kansas (EUA), e é atualmente o principal algoritmo utilizado na compressão de imagens com perdas. Ela é empregada em diversos meios multimídia, como imagens digitais (JPEG, HEIF), vídeo digital (MPEG, H.26x), áudio digital (MP3, AAC) e televisão digital (SDTV, HDTV), entre outros (SALOMON, 2007).

Segundo a decomposição em série de Fourier, funções periódicas podem ser representadas como combinações de senos e cossenos. A presença de descontinuidades, contudo, resulta em uma convergência mais lenta da série, demandando um maior número de termos para uma representação precisa. Funções mais suaves, por sua vez, apresentam melhor compactação espectral. Esse conceito fundamenta técnicas de análise no domínio da frequência e é explorado pela DCT, que utiliza apenas funções cosseno e uma extensão par do sinal, favorecendo a continuidade e a concentração de energia nos coeficientes de baixa frequência (BRITANAK; YIP; RAO, 2010).

Seja uma imagem digital, $f(x, y)$, com 8 bits de profundidade, $x = 0, 1, \ldots, 7$ e $y = 0, 1, \ldots, 7$. Segundo Rao e Yip (1990), a Transformada Discreta de Cosseno $C(u, v)$ para duas dimensões, com $u = 0, 1, \ldots, 7$ e $v = 0, 1, \ldots, 7$, é expressa por:

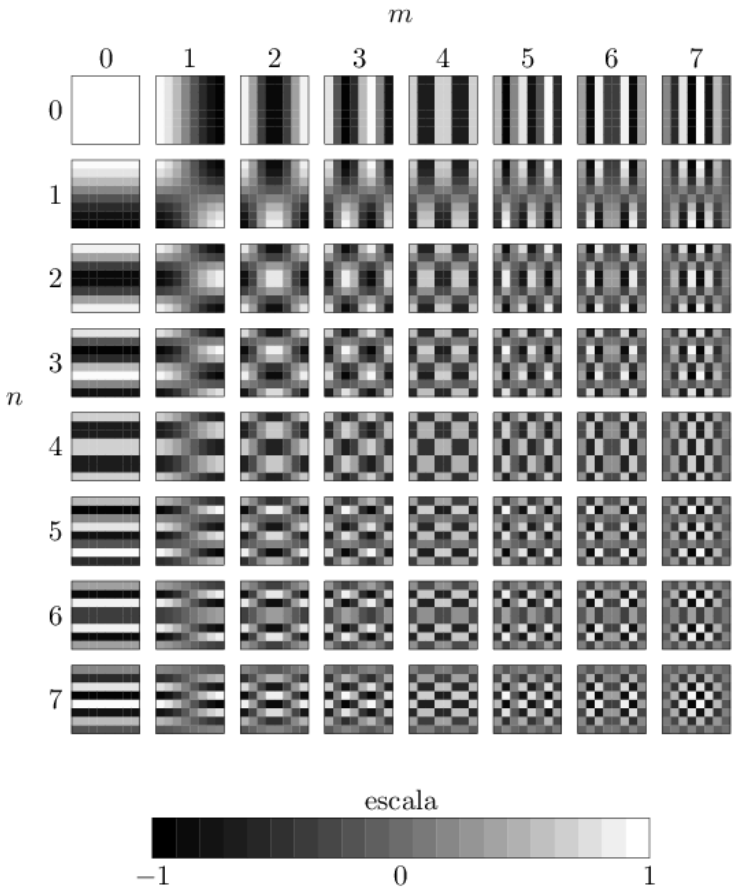
$$C(u,v) = g(u)g(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x,y) \cos\left(\frac{(2x+1)u\pi}{16}\right) \cos\left(\frac{(2y+1)v\pi}{16}\right), \tag{2.3}$$

onde:

$$g(k) = \begin{cases} \frac{1}{2}\frac{1}{\sqrt{2}} & \text{se } k = 0, \\ \frac{1}{2} & \text{se } k \neq 0. \end{cases} \tag{2.4}$$

Essencialmente, a DCT transforma as informações do domínio espacial (amplitudes) para o domínio das frequências. Por meio dela, a informação presente em cada subimagem resultante do particionamento da imagem original passa a ser representada pelos coeficientes dos 64 blocos de interpolações de cossenos, também denominados funções-base ou primitivas da DCT (STOLFI, 2016). A posição do bloco é determinada pela frequência utilizada para gerar seu conteúdo espectral; logo, quanto maior o índice do bloco, maior será a frequência. Em termos gerais, o somatório da ponderação de

Figura 5 – A base ortogonal dos 64 vetores $v_{n,m}$ da DCT bidimensional utilizados no JPEG.



Fonte: Extraído de Frasson (2013).

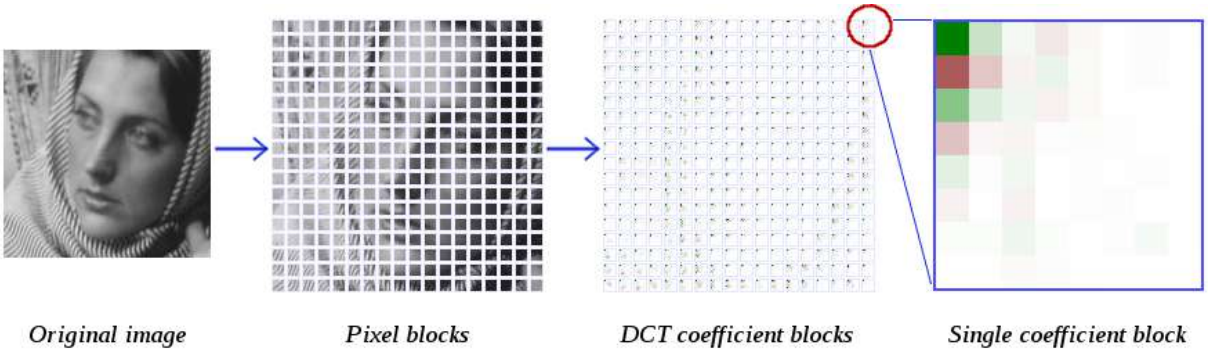
cada uma das 64 funções-base da DCT resulta no conteúdo espectral da subimagem original. A Figura 5 exibe as funções-base da DCT utilizadas na representação das subimagens.

As interpolações de cossenos apresentadas na Figura 5 representam as diferentes frequências que compõem uma imagem. É possível interpretar a frequência de uma imagem como a taxa de mudança de sua amplitude: áreas da imagem que apresentam alto contraste entre as cores, ou seja, cujas amplitudes variam abruptamente, são representadas por altas frequências; por outro lado, as áreas que exibem variações mais suaves e graduais são representadas por baixas frequências.

Ao se obter a matriz DCT de uma subimagem qualquer, é possível observar que os coeficientes na origem da DCT, localizados no canto superior esquerdo, são significativamente maiores do que àqueles com índices mais elevados, evidenciando que as frequências mais baixas constituem a maior parte da informação contida em uma imagem. Desta informação, infere-se que as altas frequências são menos importantes para a representação de uma imagem e, portanto, sua remoção (truncamento) ou diminuição de valor não alteraria significativamente o resultado final da imagem, além

de gerar uma grande economia de espaço para o armazenamento das informações. Este é o princípio utilizado para a compressão de imagens no domínio das frequências. A Figura 6 ilustra a distribuição de energia dos coeficientes DCT de um dos blocos de 8x8 píxeis uma imagem.

Figura 6 – Ilustração da conversão DCT de um bloco de 8x8 píxeis de uma imagem. A maior parte da energia dos píxeis originais se concentra nos coeficientes de baixas frequências.



Fonte: Extraído de Montgomery (2013).

Justamente por este motivo, a DCT é amplamente utilizada em algoritmos de compressão de imagens com perdas, como o padrão JPEG, no qual os coeficientes de alta frequência são quantizados de forma mais intensa ou até descartados. Esse procedimento reduz significativamente o tamanho do arquivo, preservando, contudo, uma qualidade visual aceitável.

Nesse contexto, empregar o mesmo princípio no domínio da frequência para o aprimoramento funciona como uma “inversão” do processo de degradação: ao operar diretamente sobre os coeficientes DCT, busca-se recompor detalhes e mitigar artefatos introduzidos pela compressão — não de forma estritamente reversível, mas explorando as mesmas bases e relações matemáticas utilizadas na compactação para melhorar a qualidade da imagem sob perspectivas estatísticas e perceptuais.

2.4.1 Compressão de Imagens por meio da DCT

A compressão com perdas baseada na DCT explora tanto a redundância espacial das imagens quanto as propriedades psicovisuais do sistema visual humano. Nesse processo, qualquer alteração nos coeficientes da DCT que reduza sua precisão resulta em perda de informação. Nesta subseção, apresentam-se os principais métodos empregados para a compressão de imagens no domínio da Transformada Discreta de Cosseno.

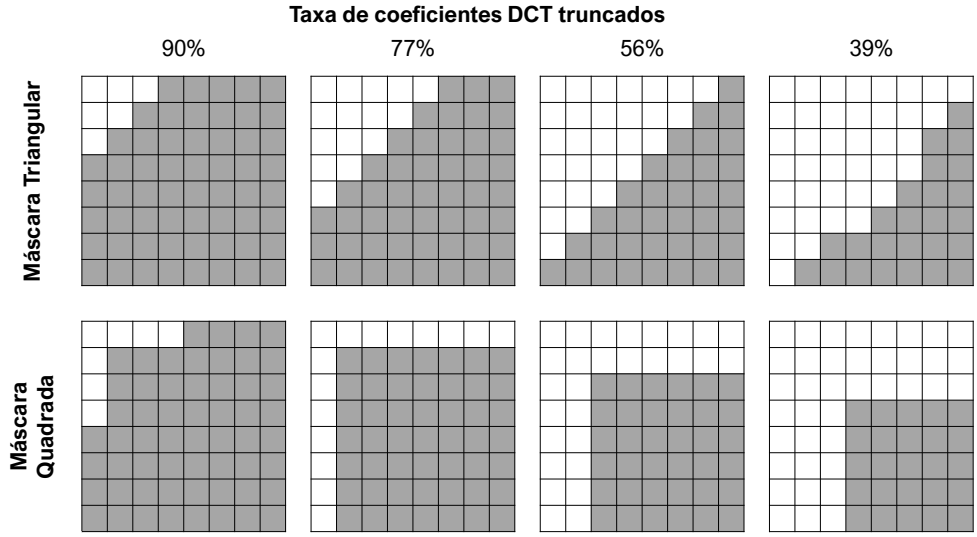
2.4.1.1 Mascaramento

O mascaramento, também conhecido como *DCT Coefficient Dropping*, consiste no descarte sistemático de coeficientes de alta frequência da DCT, aos quais o olho humano tende a ser menos sensível. Entre os métodos mais utilizados, destacam-se:

- **Mascaramento quadrado:** os coeficientes truncados formam uma região quadrada na matriz DCT. Quanto maior o quadrado de zeros, maior a taxa de compressão obtida.
- **Mascaramento triangular:** os coeficientes truncados ocupam uma região triangular inferior, alinhada à diagonal secundária. Assim como no caso quadrado, quanto maior a ordem do triângulo, maior a compressão.

A Figura 7 ilustra o mascaramento de coeficientes DCT.

Figura 7 – Mascaramentos triangular e quadrado. Os elementos brancos representam os coeficientes DCT preservados; os elementos em cinza indicam coeficientes truncados.



Fonte: De autoria própria.

A metodologia de mascaramento costumava ser empregada em meados da década de 90, sobretudo em aplicações de compressão de imagens e vídeos, quando os recursos computacionais eram limitados e a eficiência na compressão era crucial (BENYAMINOVICH; HADAR; KAMINSKY, 2005). Atualmente, o mascaramento é menos comum em aplicações modernas de compressão, mas ainda pode ser utilizado em contextos específicos onde a simplicidade e a rapidez são prioritárias, ou em sistemas embarcados com recursos limitados.

2.4.1.2 Tabelas de Quantização

Na compressão com perdas por tabelas de quantização, os coeficientes DCT são mapeados para um conjunto finito de níveis a partir de uma tabela de quantização. Uma tabela de quantização típica, $Q(u, v)$, é apresentada a seguir (ITU-T, 1992).

$$Q(u,v) = \begin{bmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{bmatrix} \tag{2.5}$$

Nesse procedimento, os coeficientes $C(u, v)$ são divididos, elemento por elemento, pelos respectivos valores da tabela de quantização, que foram multiplicados por um fator k. O resultado dessa divisão é posteriormente arredondado, resultando nos coeficientes quantizados:

$$D(u,v) = \text{round}\left(\frac{C(u,v)}{k\,Q(u,v)}\right), \tag{2.6}$$

onde k é o fator que escala a intensidade da quantização.

O valor de k é determinado a partir do parâmetro de qualidade q. Para $1 \leq q \leq 99$, tem-se:

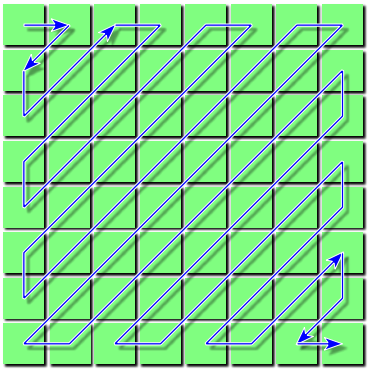
$$k(q) = \begin{cases} \dfrac{5000}{q}, & \text{se } 1 \leq q \leq 50, \\ 200 - 2q, & \text{se } 51 \leq q \leq 99. \end{cases} \tag{2.7}$$

Todos esses processos encontram-se representados no Bloco 2 da Figura 1. As tabelas de quantização padrão do JPEG estão dispostas no Apêndice A.

Por fim, os coeficientes são vetorizados em uma sequência zigzag (Figura 8), cujo objetivo é agrupar coeficientes nulos nas últimas posições, facilitando a posterior codificação por entropia.

A codificação por entropia atribui códigos de comprimento variável aos símbolos de acordo com suas probabilidades de ocorrência: símbolos mais frequentes recebem códigos mais curtos, enquanto símbolos menos frequentes recebem códigos mais longos, reduzindo o tamanho final do arquivo (OLIVEIRA, 2002). Como grande parte dos coeficientes quantizados é igual a zero em decorrência do processo de mascaramento e arredondamento, aplica-se inicialmente o RLE (*Run-Length Encoding*), que é uma técnica de compressão simples e eficaz para sequências com muitos valores repetidos. O

Figura 8 – Conversão de matriz (8,8) para vetor unidimensional (1D).



Fonte: Khristov (2007).

RLE substitui sequências consecutivas de valores idênticos por um único valor seguido do número de repetições, reduzindo significativamente o tamanho dos dados quando há longas sequências de zeros, como é comum após a quantização na compressão JPEG (Stanford University, n.d.). Esse processo é seguido pela codificação de Huffman, que é um algoritmo de compressão sem perdas que utiliza códigos de comprimento variável para representar os símbolos com base em suas frequências de ocorrência (HUFFMAN, 1952). A combinação dessas técnicas resulta em uma compressão sem perdas dos dados quantizados, reduzindo significativamente o tamanho do arquivo final. Ambos os procedimentos são representados no Bloco 3 da Figura 1.

2.4.2 Implicações da compressão DCT na qualidade da imagem

A compressão de imagens utilizando a DCT pode introduzir diversos artefatos visuais que afetam a qualidade percebida da imagem. Esses artefatos resultam das perdas de informação inerentes ao processo de quantização e mascaramento dos coeficientes DCT. Entre os principais artefatos causados pela compressão DCT, destacam-se:

2.4.2.1 Blocos visíveis

Um dos artefatos mais comuns é a aparição de blocos visíveis na imagem comprimida. Isso ocorre quando os coeficientes DCT de alta frequência são severamente quantizados ou descartados, resultando em transições abruptas entre os blocos de 8×8 píxeis. Esses blocos podem ser percebidos como linhas ou bordas artificiais, especialmente em áreas de baixa textura ou cor uniforme, como pode ser visto na Figura 9.

Figura 9 – Surgimento de blocos visíveis em imagem comprimida com perdas.

(a) Imagem original

(b) Imagem JPEG comprimida com $Q = 10$

Fonte: De autoria própria.

2.4.2.2 *Ringing*

O artefato de *ringing* manifesta-se como halos ou sombras ao redor de bordas nítidas na imagem. Isso ocorre devido à perda de coeficientes DCT que representam altas frequências, o que pode causar oscilações indesejadas na reconstrução da imagem, especialmente em regiões de alto contraste.

Figura 10 – Surgimento de *ringing* em imagem comprimida com perdas.

(a) Imagem original

(b) Imagem comprimida

Fonte: Wikipedia contributors (2025).

2.4.2.3 *Mosquito Noise*

O *mosquito noise* é caracterizado por pequenas manchas ou pontos que surgem em torno de áreas de alto contraste, como bordas ou detalhes finos. Esse artefato resulta da quantização dos coeficientes DCT, a qual pode introduzir ruído visual na imagem comprimida. A Figura 11 exibe um comparativo entre uma imagem de alta fidelidade e sua versão com alta taxa de compressão, a qual resultou no surgimento do *mosquito noise*.

Figura 11 – Surgimento de *mosquito noise* em imagem comprimida com perdas. As setas brancas indicam o *mosquito noise*.

(a) Imagem original (b) Imagem degradada

Fonte: Algolith (2025).

Com o entendimento do funcionamento da DCT e de suas implicações na qualidade da imagem, torna-se necessária a compreensão de como a inteligência artificial pode auxiliar na mitigação dos danos causados a imagens degradadas. A seguir, apresentam-se os fundamentos teóricos de Redes Neurais Artificiais que embasam o desenvolvimento deste trabalho.

2.5 REDES NEURAIS ARTIFICIAIS

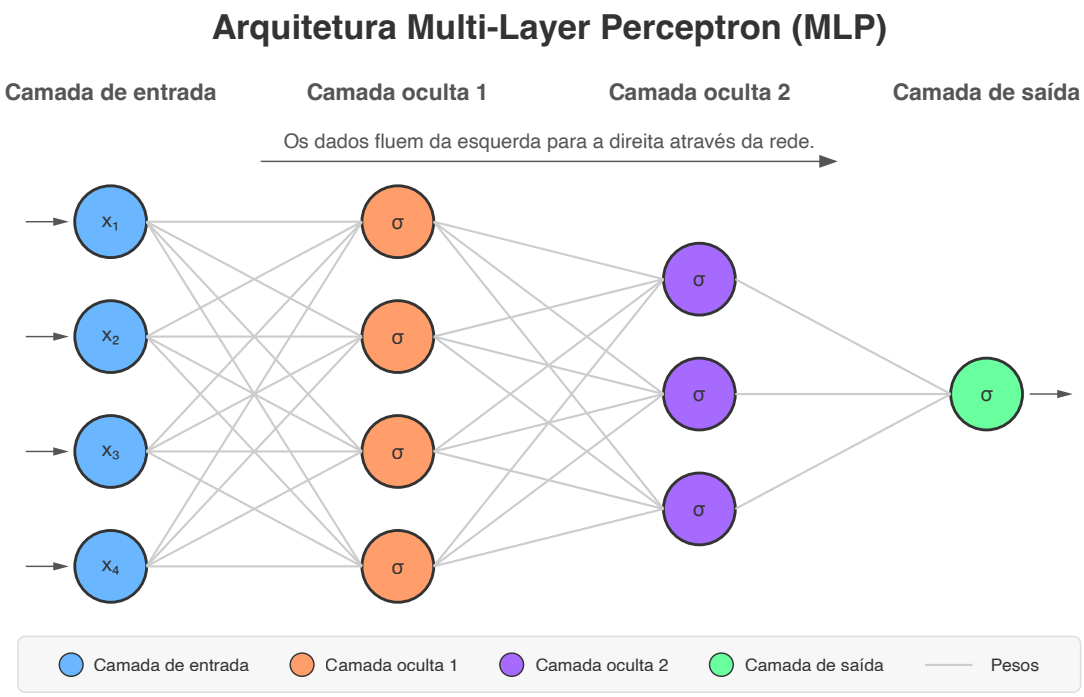
Segundo Haykin (2009), as redes neurais artificiais são sistemas computacionais inspirados na organização e no funcionamento do cérebro humano. Em sua definição clássica, uma rede neural é:

> "Um processador paralelo massivamente distribuído constituído por unidades de processamento simples, que têm uma propensão natural para armazenar conhecimento experimental e torná-lo disponível para uso. Ela se assemelha ao cérebro em dois aspectos: (i) o conhecimento é adquirido pela rede a partir de seu ambiente por meio de um processo de aprendizado; e (ii) as forças de conexão entre os neurônios, conhecidas como pesos sinápticos, são utilizadas para armazenar o conhecimento adquirido."

As redes neurais modernas são compostas por camadas de neurônios artificiais que realizam combinações lineares dos dados de entrada, seguidas por operações não lineares, o que permite que o modelo aprenda funções complexas sem a necessidade de programação explícita. Um exemplo de Rede Neural Artificial é apresentado pelo diagrama da Figura 12.

No contexto do aprendizado supervisionado adotado neste trabalho, a rede é treinada a partir de pares entrada-saída, aprendendo uma transformação que se

Figura 12 – Exemplo de Rede Neural Artificial do tipo MLP, com múltiplas camadas de neurônios conectadas.



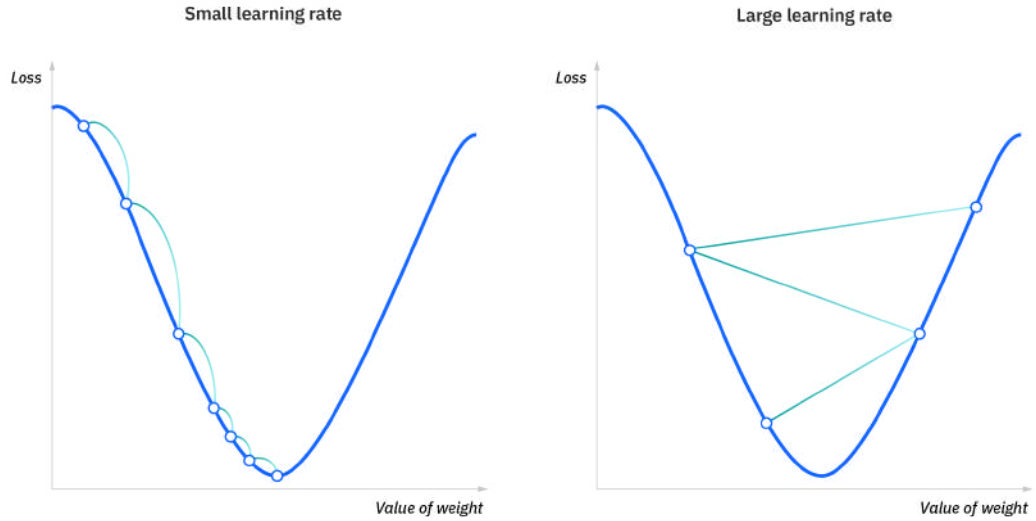
Fonte: Adaptação de CodeSignal (n.d.).

aproxima de um mapeamento desejado. Uma época de treinamento corresponde a uma passagem completa pelo conjunto de dados de treinamento, durante a qual os parâmetros do modelo são atualizados de forma iterativa. Esse processo é repetido ao longo de múltiplas épocas, até que a rede convirja para uma solução capaz de generalizar adequadamente para dados não apresentados durante o treinamento.

Durante o processo de treinamento, os parâmetros dessas camadas — pesos e vieses — são ajustados por meio de algoritmos de otimização, tipicamente variantes do Gradiente Descendente, com o objetivo de minimizar uma função de perda que quantifica a discrepância entre a saída predita pela rede e a saída desejada. A Figura 13 ilustra, de forma comparativa, o comportamento do gradiente descendente sob taxas de aprendizado pequenas e grandes, destacando as diferenças na trajetória e na estabilidade do processo de otimização.

Analisando a Figura 13, observa-se que uma taxa de aprendizado pequena resulta em uma trajetória de otimização mais estável, porém com uma convergência mais lenta, exigindo um maior número de iterações (épocas) para alcançar o mínimo da função de perda. Por outro lado, uma taxa de aprendizado grande pode acelerar o processo de convergência, mas também pode levar a oscilações e instabilidades, dificultando a aproximação do mínimo global e potencialmente resultando em divergência. Isso evidencia a importância de um ajuste cuidadoso da taxa de aprendizado para

Figura 13 – Comportamento do gradiente descendente com taxa de aprendizado pequena (esq.) e grande (dir.).



Fonte: Adaptado de IBM (2025).

equilibrar velocidade e estabilidade durante o treinamento da rede neural.

2.5.1 Redes MLP e funções de ativação

Uma arquitetura amplamente utilizada em Redes Neurais Artificiais (RNA) é a *Multilayer Perceptron* (MLP), que é composta por camadas densamente conectadas (*fully connected*). Cada neurônio artificial realiza a seguinte operação:

$$y = \varphi(\mathbf{w}^\top \mathbf{x} + b), \tag{2.8}$$

em que $\mathbf{x}$ é o vetor de entrada, $\mathbf{w}$ é o vetor de pesos, b é o viés e $\varphi(\cdot)$ é uma função de ativação não linear. O uso de funções não lineares, como ReLU (AGARAP, 2018), SiLU (ELFWING; UCHIBE; DOYA, 2018) ou GELU (HENDRYCKS; GIMPEL, 2023), torna a rede capaz de modelar relações complexas nos dados, que não poderiam ser representadas por modelos puramente lineares.

Neste trabalho, funções de ativação mais recentes, como a GELU (*Gaussian Error Linear Unit*), foram utilizadas devido à sua suavidade e boa estabilidade numérica, características vantajosas para tarefas de regressão de alta dimensionalidade, como a predição de coeficientes DCT.

2.5.2 Normalização e estabilização do treinamento

Com o objetivo de melhorar a estabilidade do treinamento e acelerar a convergência do modelo, foi empregada a técnica de normalização de camadas (*Layer*

Normalization) (XU et al., 2019). Essa técnica normaliza as ativações ao longo das dimensões de características de cada amostra, tornando o treinamento menos sensível à escala das ativações e mais estável em redes profundas. Como consequência, a técnica contribui indiretamente para uma melhor capacidade de generalização.

A normalização imposta pela *Layer Normalization* força as ativações a apresentarem média nula e variância unitária, o que auxilia na mitigação de problemas como a explosão ou a dissipação dos gradientes, frequentemente observados em redes neurais com múltiplas camadas intermediárias.

2.5.3 Arquiteturas residuais

Uma característica central das redes modernas é o uso de *skip-connections* ou conexões residuais, introduzidas inicialmente em He et al. (2016), que originou o termo *Residual Networks*, ou *ResNet*. Nessa abordagem, a camada não aprende uma transformação completa, mas sim um *resíduo* a ser somado à entrada original:

$$\mathbf{y} = \mathbf{x} + \alpha \cdot \mathcal{F}(\mathbf{x}), \tag{2.9}$$

em que $\mathcal{F}(\cdot)$ representa a transformação aprendida pela rede e α é um fator de escala residual.

Essa estratégia é particularmente relevante neste trabalho, pois os modelos utilizados realizam a predição de correções nos coeficientes DCT, e não a reconstrução absoluta desses coeficientes. A natureza residual da tarefa reflete a hipótese de que a imagem degradada já contém grande parte da estrutura da imagem original, sendo necessário apenas corrigir artefatos e recuperar detalhes perdidos na compressão JPEG. A arquitetura desenvolvida neste trabalho, denominada `ContextResidualMLP`, combina informações do bloco central e de sua vizinhança, demonstrando-se vantajosa para a captura de dependências locais no domínio da frequência.

2.5.4 Funções de perda e aprendizado no domínio da Frequência

O treinamento das redes neurais neste trabalho envolve funções de perda que operam diretamente sobre os coeficientes da DCT, possibilitando a atuação no domínio da frequência. Funções tradicionais, como o Erro Quadrático Médio (MSE do inglês Mean Squared Error) e o Erro Absoluto Médio – que iremos denotar aqui por L1, por ser considerada como a média da norma L1, foram avaliadas juntamente com o cálculo de perdas ponderadas por frequência, que atribui maior importância a coeficientes de alta frequência — aquelas que são mais sensíveis à quantização no padrão JPEG.

A combinação de perdas, como a MSE e a perda ponderada em frequência, apresentou resultados satisfatórios, principalmente por equilibrar a reconstrução global e a recuperação de detalhes finos.

2.6 MÉTRICAS DE AVALIAÇÃO

Para que seja possível analisar os resultados de maneira quantitativa, métricas de avaliação são necessárias. Neste trabalho, utilizamos duas métricas *Full-Reference* (FR), que comparam a imagem melhorada com a imagem original de alta qualidade; e duas métricas *No-Reference* (NR), que avaliam a qualidade da imagem melhorada sem a necessidade da imagem original. As métricas FR utilizadas foram: Peak Signal-to-Noise Ratio (PSNR) e Multiscale Structural Similarity (MS-SSIM). Já as métricas NR empregadas foram: Natural Image Quality Evaluator (NIQE) e entropia. A seguir, detalhamos as métricas utilizadas.

2.6.1 Métricas de referência completa

As métricas de referência completa comparam a imagem processada à imagem original, considerada referência ideal. Essas métricas são amplamente utilizadas para quantificar a qualidade de imagens restauradas ou aprimoradas, fornecendo uma medida objetiva da fidelidade em relação à imagem original.

O Erro Quadrático Médio (MSE - *Mean Square Error*), é uma das métricas mais simples e amplamente utilizadas para medir a diferença entre dois sinais. O MSE fundamenta diversas outras métricas de avaliação, como a PSNR, utilizada neste trabalho. É definido como a média da diferença entre o sinal ideal e o sinal obtido, elevada ao quadrado (GONZALEZ; WOODS, 2018).

Sejam $f(n)$ e $g(n)$ representações do valor (intensidade) de um pixel em uma imagem na localização $n = (n1, n2)$. O MSE entre as duas representações é definido como:

$$\mathrm{MSE}\left[f(n), g(n)\right] = \frac{1}{N} \sum_{n} \left[f(n) - g(n)\right]^2, \tag{2.10}$$

sendo N o número total de píxeis da imagem.

A Relação Sinal-Ruído de Pico (PSNR) define a relação entre a energia máxima de um sinal e o ruído que afeta sua representação fidedigna (BOVIK, 2009). A métrica, advinda do MSE, é expressa em uma escala logarítmica, e pode ser obtida pela seguinte equação:

$$\mathrm{PSNR}\left[f(n), g(n)\right] = 10 \log_{10} \frac{E^2}{\mathrm{MSE}\left[f(n), g(n)\right]}, \tag{2.11}$$

onde E é o valor máximo que um pixel pode assumir. Para uma imagem RGB com profundidade de cor de 8 bits, $E = 255$. Na PSNR, obtemos resultados normalizados; ou seja, o valor zero representa o pior resultado possível para a imagem comprimida, restando apenas ruído, enquanto o valor 100 indica que as imagens são idênticas (HWANG, 2009).

Apesar de ambas as métricas serem amplamente difundidas, elas não consideram aspectos psicofísicos do sistema visual humano, o que pode levar a resultados que não corroboram a percepção visual humana. Por esse motivo, foram desenvolvidas métricas mais sofisticadas, como o Índice de Similaridade Estrutural (SSIM) e o Multiscale Structural Similarity (MS-SSIM).

O SSIM é um índice utilizado para medir a similaridade entre duas imagens. Ele é comumente empregado em processamento de imagens e visão computacional para avaliar a qualidade de uma imagem, frequentemente no contexto de compressão ou restauração de imagens (GUNAWAN et al., 2025). Diferentemente das métricas tradicionais que comparam diferenças pixel a pixel, o SSIM avalia a qualidade percebida de uma imagem considerando as alterações na informação estrutural, luminância e contraste (WANG et al., 2004).

Em sua formulação geral, o SSIM é definido como:

$$\mathrm{SSIM}_{(t,r)} = [s_{(t,r)}]^{\alpha} \cdot [l_{(t,r)}]^{\beta} \cdot [c_{(t,r)}]^{\gamma}, \tag{2.12}$$

onde $s_{(t,r)}$, $l_{(t,r)}$ e $c_{(t,r)}$ são as medidas de similaridade estrutural, luminância e contraste, respectivamente, entre as imagens de teste t e de referência r. Os expoentes α, β e γ são parâmetros de ponderação que ajustam a importância relativa de cada componente na métrica final.

Com as escolhas usuais $\alpha = \beta = \gamma = 1$ e as constantes de estabilização positivas C_1 e C_2, a forma amplamente utilizada do SSIM é:

$$\mathrm{SSIM}(t,r) = \frac{(2\mu_t\mu_r + C_1)(2\sigma_{tr} + C_2)}{(\mu_t^2 + \mu_r^2 + C_1)(\sigma_t^2 + \sigma_r^2 + C_2)}, \tag{2.13}$$

onde μ_t e μ_r são as médias locais das imagens de teste e referência, σ_t^2 e σ_r^2 são suas variâncias locais, e σ_{tr} é a covariância local entre elas. O valor do SSIM é igual a 1 quando há similaridade perfeita; na prática, com $C_1, C_2 > 0$, os valores observados situam-se tipicamente no intervalo $[0, 1]$ para imagens que representam cenas reais do mundo físico, capturadas por câmeras.

O *Multi-Scale SSIM* (MS-SSIM) é uma extensão do SSIM, projetada para capturar a percepção visual humana de maneira mais abrangente, considerando que o olho humano percebe detalhes em diferentes níveis de resolução (WANG; SIMONCELLI; BOVIK, 2003). Nesse método, as imagens são submetidas a um processo de pirâmide gaussiana, no qual são obtidas múltiplas escalas por meio de suavização e *downsampling* sucessivos. Em cada escala, calculam-se os termos de contraste $c_j(t,r)$ e estrutura $s_j(t,r)$, enquanto a luminância $l_M(t,r)$ é avaliada apenas na escala mais grossa. A métrica final é obtida pela combinação ponderada dos valores em todas as

escalas, dada por:

$$\mathrm{MS-SSIM}_{(t,r)} = [l_M(t,r)]^{\alpha_M} \prod_{j=1}^{M} [c_j(t,r)]^{\beta_j} [s_j(t,r)]^{\gamma_j}, \tag{2.14}$$

onde M representa o número total de escalas e os expoentes α_M, β_j e γ_j controlam a contribuição relativa de cada componente em cada nível de resolução.

Estudos demonstram que o MS-SSIM apresenta maior correlação com a percepção visual humana em relação ao SSIM de única escala, especialmente em cenários de compressão e restauração de imagens (WANG; SIMONCELLI; BOVIK, 2003).

2.6.2 Métricas sem referência

As métricas sem referência avaliam a qualidade de uma imagem sem a necessidade da imagem original de alta qualidade. Neste trabalho, utilizam-se duas métricas desse tipo: o *Natural Image Quality Evaluator* (NIQE) e a entropia.

O NIQE é uma métrica de qualidade de imagem *sem referência* proposta por Mittal, Soundararajan e Bovik (2013), que avalia a qualidade perceptual de uma imagem com base nas estatísticas de cenas naturais (*Natural Scene Statistics* — NSS). A métrica é calculada pela distância entre dois modelos Gaussianos multivariados (MVG): um ajustado às imagens de paisagens naturais (prístinas) e outro ajustado à imagem analisada. Essa distância é dada por:

$$\mathrm{NIQE}(I) = \sqrt{(\nu_1 - \nu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1} (\nu_1 - \nu_2)}, \tag{2.15}$$

onde ν_1 e ν_2 são os vetores de médias, e Σ_1 e Σ_2 são as matrizes de covariância dos modelos Gaussianos correspondentes ao conjunto natural e à imagem testada, respectivamente (MITTAL; SOUNDARARAJAN; BOVIK, 2013). Valores mais baixos de NIQE indicam melhor qualidade de imagem, enquanto valores mais altos correspondem a maiores desvios das estatísticas naturais, ou seja, pior qualidade perceptual.

A entropia é uma medida da quantidade de informação contida em uma imagem. Em termos simples, a entropia quantifica a incerteza ou imprevisibilidade associada aos valores dos píxeis de uma imagem (SHANNON, 1948). Ela é calculada com base na distribuição de probabilidade $p(i)$ dos níveis de cinza presentes na imagem, conforme a seguinte equação:

$$H = -\sum_i p(i) \log_2 p(i), \tag{2.16}$$

onde a soma é realizada sobre todos os valores possíveis de intensidade. A entropia é medida em bits/símbolo; valores mais altos indicam maior diversidade de informação,

enquanto valores baixos refletem regiões mais homogêneas e previsíveis. Em geral, imagens com maior entropia podem ser consideradas mais ricas em detalhes e texturas, enquanto imagens com baixa entropia tendem a parecer mais simples ou suaves. Contudo, a entropia, por si só, não é suficiente para avaliar a qualidade perceptual de uma imagem, sendo mais eficaz quando combinada com outras métricas.

3 METODOLOGIA

A elaboração deste trabalho teve início com o estudo dos fundamentos teóricos apresentados no capítulo anterior, tendo como objetivo a compreensão dos procedimentos envolvidos no processamento de imagens e na construção de modelos de aprendizado de máquina (ML - *Machine Learning*). A seguir, detalha-se a metodologia adotada, incluindo a aquisição e preparação dos dados, a arquitetura da rede neural proposta, o procedimento de treinamento e as ferramentas utilizadas.

3.1 VISÃO GERAL

A metodologia desenvolvida neste trabalho abrange seis etapas principais:

(i) aquisição do conjunto de dados e divisão das imagens em conjuntos de treinamento, validação e teste;

(ii) degradação controlada por compressão JPEG no domínio da DCT;

(iii) extração dos coeficientes no domínio da DCT;

(iv) construção de um conjunto de dados estruturado em arquivo HDF5;

(v) treinamento supervisionado da rede;

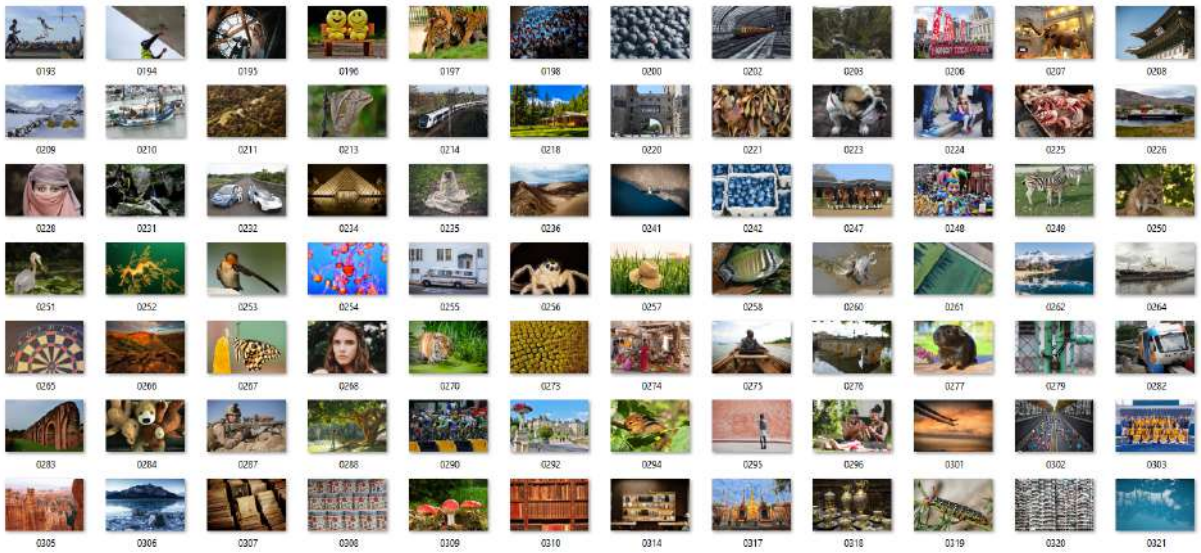
(vi) avaliação quantitativa e visual nas imagens de teste.

O Apêndice A apresenta um diagrama detalhado do fluxo de pré-processamento dos dados, ilustrando as interações entre as etapas (i), (ii), (iii) e (iv). O Apêndice B apresenta um diagrama que compreende as etapas (v) e (vi) da metodologia desenvolvida neste trabalho. Todas as etapas são discutidas em detalhes nas seções subsequentes.

3.2 AQUISIÇÃO DOS DADOS

Considerando a complexidade inerente à programação e às ferramentas requeridas, iniciou-se o desenvolvimento prático do projeto com a etapa de aquisição dos dados — isto é, a seleção do *dataset* de imagens a ser utilizado no treinamento e nos testes dos modelos de redes neurais. Após análise criteriosa de trabalhos acadêmicos relacionados e consulta a repositórios de dados abertos, optou-se pelo uso do *dataset* DIV2K (AGUSTSSON; TIMOFTE, 2017), composto por 900 imagens RGB em alta resolução. Tal conjunto de dados tem sido amplamente adotado em trabalhos na área de aprimoramento de imagens (*image enhancement*) em razão de sua qualidade e diversidade visual. A Figura 14 apresenta parte das imagens que compuseram a base de dados.

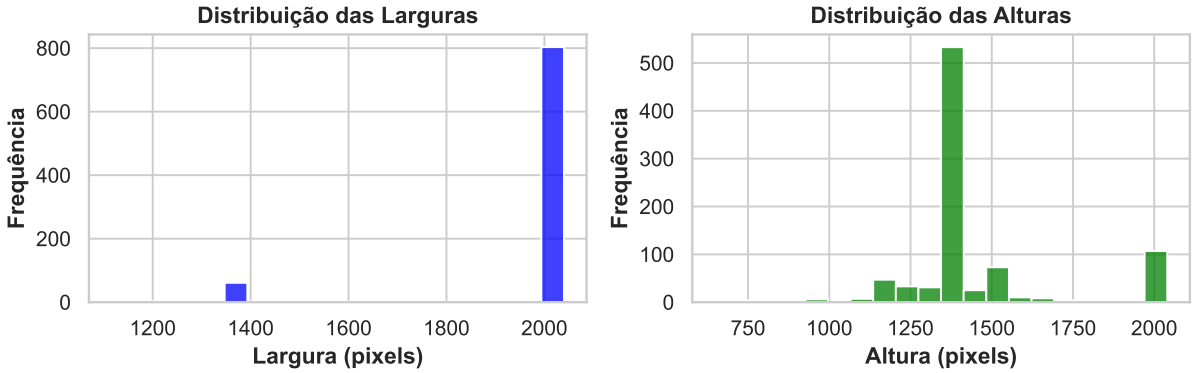
Figura 14 – Exemplos de imagens do *dataset* DIV2K utilizadas para treinamento e avaliação dos modelos propostos.



Fonte: De autoria própria.

A fim de padronizar o processamento das imagens para o treinamento dos modelos, definiu-se que todas as imagens deveriam apresentar a mesma resolução, cujos valores fossem múltiplos de 8, de modo a facilitar a aplicação da DCT. Para identificar a resolução mais recorrente e, assim, definir-se a resolução a ser adotada, analisou-se a distribuição das resoluções presentes no DIV2K, conforme ilustra a Figura 15.

Figura 15 – Distribuição das dimensões das imagens do *dataset* DIV2K.



Fonte: De autoria própria.

Conforme os gráficos da Figura 15 indicaram, observou-se que uma parcela significativa das imagens possuía a resolução de 2040 $\times$ 1356 píxeis, representando a moda do conjunto de dados. Dessa forma, convencionou-se que as imagens deveriam possuir o formato paisagem. Considerando o requisito de múltiplos exatos de 8 e as restrições de processamento e treinamento dos modelos, realizou-se o *center-crop* das imagens para a razão de aspecto 3:2, seguido de redimensionamento utilizando o filtro Lanczos, que proporcionou boa preservação das altas frequências. Assim, as imagens

com resolução original de 2040 $\times$ 1356 píxeis foram recortadas e redimensionadas para 1200 $\times$ 800 píxeis. Imagens com outras resoluções foram tratadas de forma análoga, assegurando a padronização de todo o conjunto de dados.

3.3 DEGRADAÇÃO CONTROLADA

Com o objetivo de reproduzir de forma controlada os efeitos de degradação comumente observados em imagens comprimidas, e permitir que a rede neural aprendesse a restaurar detalhes perdidos ou distorcidos, adotou-se um procedimento supervisionado de geração de pares *entrada–referência*. Nessa abordagem, cada imagem degradada (entrada do modelo) foi derivada de uma imagem de alta qualidade (imagem de referência), de modo que a rede pudesse reconhecer padrões na degradação e, assim, aprender a mapear a versão degradada à sua correspondente de alta qualidade.

Diferentemente dos métodos de degradação baseados em reamostragem espacial (*downsampling* e *upsampling*), que introduzem perda de informação principalmente por suavização e interpolação, neste trabalho a degradação foi realizada por meio da compressão JPEG, aproximando o processo de degradação de cenários reais e estabelecendo uma coerência teórica com o domínio de atuação do modelo.

Para a geração das versões degradadas, definiu-se a utilização de cinco níveis Q de qualidade na compressão JPEG: $Q \in \{10, 20, 30, 40, 50\}$. Cada imagem foi comprimida com apenas um desses níveis de qualidade, de forma que o conjunto de teste foi estratificado com 20% das imagens para cada valor de Q. Dessa forma, o conjunto de testes abrange desde condições de alta compressão até níveis moderados de compressão, representando um cenário realista de uso do formato JPEG.

O processo de compressão foi automatizado utilizando a biblioteca `Pillow`, que permitiu a definição explícita do parâmetro de qualidade JPEG. As imagens resultantes foram armazenadas em diretórios separados conforme o valor de Q, o que facilitou a organização e o acesso durante as etapas subsequentes de extração dos coeficientes DCT e treinamento dos modelos.

3.4 EXTRAÇÃO DOS COEFICIENTES DCT

Inicialmente, cada imagem foi convertida do espaço de cor RGB para YCbCr, separando os canais de luminância (Y) e crominância (Cb, Cr), conforme a Equação 2.1. Em seguida, cada canal de cor foi particionado em blocos 8×8, sobre os quais a DCT foi aplicada, segundo a Equação 2.3. Os 64 coeficientes (u, v) foram então vetorizados via percurso zigue-zague, consolidando um vetor de dimensão 64 por bloco — conforme ilustrado na Figura 8. A extração foi realizada tanto para as imagens de referência (alta

qualidade) quanto para as imagens degradadas (compressão JPEG), gerando pares de vetores (x, y) para treinamento supervisionado.

3.4.1 Normalização e Tipo Numérico

Para estabilizar o aprendizado e manter a coerência entre os canais de cor, normalizou-se cada vetor $c \in \mathbb{R}^{64}$ no intervalo aproximado $[-1, 1]$ por meio de uma divisão escalar com um fator de normalização constante $S = 2040$. O valor foi definido com base na análise teórica da DCT-2D, correspondente ao máximo teórico do coeficiente DC para um bloco 8×8 saturado (todos os píxeis com valor 255).

A DCT bidimensional é definida pela seguinte fórmula:

$$\text{DCT}_{u,v} = \frac{1}{4} \cdot C(u) \cdot C(v) \cdot \sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y) \cdot \cos\left[\frac{(2x+1)u\pi}{16}\right] \cdot \cos\left[\frac{(2y+1)v\pi}{16}\right].$$

Para o coeficiente DC, tem-se $u = 0$ e $v = 0$, o que simplifica a expressão para:

$$\text{DCT}_{0,0} = \frac{1}{4} \cdot C(0)^2 \cdot \sum_{x=0}^{7} \sum_{y=0}^{7} 255.$$

Sabendo que $C(0) = \frac{1}{\sqrt{2}}$, e que a soma possui 64 termos:

$$\text{DCT}_{0,0} = \frac{1}{4} \cdot \left(\frac{1}{\sqrt{2}}\right)^2 \cdot 64 \cdot 255 = 8 \cdot 255 = \boxed{2040}. \tag{3.1}$$

Logo, o coeficiente DC máximo teórico é de 2040, justificando a escolha do fator de normalização $S = 2040$ e assegurando que os coeficientes normalizados permanecessem dentro do intervalo desejado.

3.5 CONSTRUÇÃO DO DATASET EM HDF5

3.5.1 Geração Paralela dos Coeficientes DCT

A geração dos coeficientes foi paralelizada com `ProcessPoolExecutor`, processando cada imagem de forma independente, segundo a sequência abaixo:

(i) leitura da imagem de referência e da imagem degradada (qualidade Q);

(ii) conversão do sistema de cores RGB para YCbCr;

(iii) particionamento em blocos 8×8 e cálculo da DCT por bloco;

(iv) vetorização zigue-zague e normalização;

(v) armazenamento dos pares (x, y) no arquivo `.h5` correspondente ao canal de cor selecionado (Y, Cb ou Cr), com dados de qualidades $Q \in \{10, 20, 30, 40, 50\}$.

O conjunto final foi salvo em um único arquivo HDF5 contendo:

- `X_train`, `Y_train`;
- `X_validation`, `Y_validation`;
- `X_test`, `Y_test`.

O formato HDF5 se destaca por sua eficiência na leitura e escrita de grandes volumes de dados, além de permitir o armazenamento estruturado em grupos e datasets, facilitando o acesso e a manipulação dos dados durante o treinamento dos modelos.

Além do arquivo H5, foram gerados arquivos de manifesto (JSON/CSV/XLSX) com o seguinte mapeamento: `split, qualidade da imagem, índice da imagem e caminho do arquivo`, a fim de facilitar a inspeção e o reuso dos dados.

3.5.2 Divisão dos Dados

A divisão dos dados seguiu a proporção $80\%/10\%/10\%$ para os conjuntos de treinamento, validação e teste, respectivamente. A separação foi realizada em nível de imagem, de modo que nenhuma imagem estivesse presente em mais de uma partição. Cada partição continha imagens de todas as qualidades JPEG definidas (de $Q = 10$ a $Q = 90$), assegurando diversidade e representatividade. A seleção das imagens em cada partição foi conduzida de forma aleatória, com utilização de uma semente determinística, garantindo a reprodutibilidade dos experimentos.

Para imagens de resolução 1200×800, cada imagem é dividida em 150×100 blocos DCT 8×8, totalizando 15000 blocos por imagem. O modelo `ContextResidualMLP` utiliza uma vizinhança 3×3 de blocos DCT, resultando em entradas de dimensão 576, ou seja, nove blocos concatenados. Essa vizinhança é extraída em regiões de borda da imagem, utilizando o modo de reflexão *reflect padding* para definir blocos vizinhos quando estes extrapolam os limites espaciais, assegurando consistência dimensional e evitando perda de informação Vemulapati (2023).

Considerando o total de 900 imagens no *dataset*, o conjunto completo gerou $900 \times 15000 = 13{,}500{,}000$ amostras. Com a divisão $80\%/10\%/10\%$, obteve-se:

- **Treinamento:** 720 imagens, resultando em $10{,}800{,}000$ amostras;
- **Validação:** 90 imagens, resultando em $1{,}350{,}000$ amostras;
- **Teste:** 90 imagens, resultando em $1{,}350{,}000$ amostras.

Cada amostra é composta pela vizinhança 3×3 de blocos DCT (entrada) e pelo bloco central correspondente (saída desejada), ambos normalizados e organizados conforme os passos descritos anteriormente.

3.6 FERRAMENTAS E AMBIENTE EXPERIMENTAL

O treinamento dos modelos foi realizado na plataforma Google Colab Pro, utilizando o ambiente com GPU A100 (80 GB de VRAM), o que auxiliou significativamente o processo de testes de diferentes arquiteturas de rede, bem como o aprendizado. Além disso, o ambiente também foi empregado para testes e validações preliminares.

Os experimentos foram conduzidos utilizando a linguagem Python 3.11, juntamente com as seguintes bibliotecas:

- PyTorch: criação, treino e validação do modelo;
- OpenCV: DCT e iDCT por bloco e operações matriciais;
- Pillow: entrada e saída de imagens, bem como conversões de sistemas de cor;
- NumPy/Pandas: manipulação de arrays e geração de arquivos manifesto;
- h5py: leitura e escrita de arquivos HDF5;
- Matplotlib/Seaborn: visualização de dados e resultados;
- tqdm: exibição de barras de progresso durante o processamento;
- scikit-image e sewar: cálculo de métricas de qualidade de imagem.

3.7 MODELOS DE REDES NEURAIS

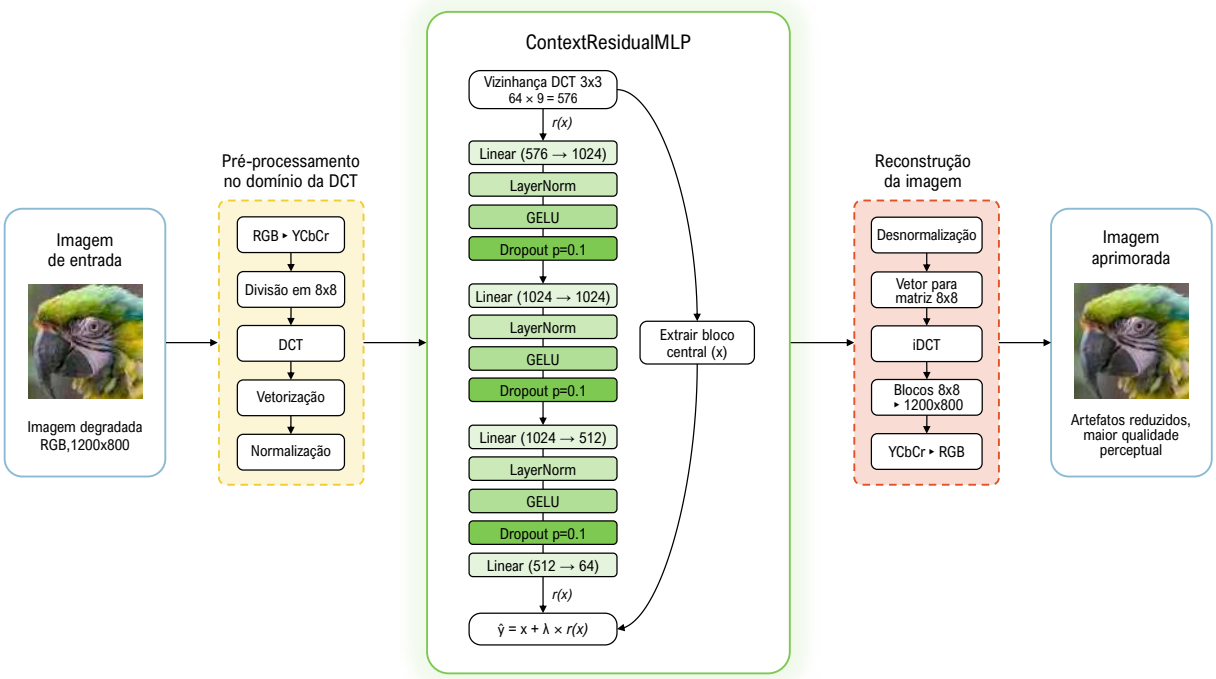
3.7.1 Arquitetura Proposta: ContextResidualMLP

A etapa de aprimoramento dos coeficientes DCT foi conduzida por uma rede neural totalmente conectada projetada para operar diretamente no domínio da frequência.

A arquitetura da rede proposta neste trabalho, denominada `ContextResidualMLP`, fundamenta-se na premissa de que a reconstrução de um bloco degradado pode ser melhor estimada quando considera não apenas seus próprios coeficientes, mas também os coeficientes presentes em sua vizinhança espacial imediata no domínio da DCT. Essa abordagem contextual permite capturar dependências locais entre blocos adjacentes, refletindo a redundância espacial observada em imagens naturais e os padrões estruturais preservados pela transformada. A arquitetura de rede proposta está ilustrada na Figura 16.

A `ContextResidualMLP` é composta por três camadas totalmente conectadas com 1024, 1024 e 512 neurônios, respectivamente. Cada camada é seguida por normalização de camada (*Layer Normalization*), ativação não linear GELU e um termo de regularização por *dropout* ($p = 0{,}10$). Esses recursos promovem a estabilidade

Figura 16 – Visão geral do fluxo de processamento para aprimoramento de imagens comprimidas em JPEG através dos modelos `ContextResidualMLP`. A Figura destaca o pré-processamento dos dados, a arquitetura da rede neural e o pós-processamento para reconstrução da imagem final.



Fonte: De autoria própria.

numérica, a capacidade de modelagem não linear e uma maior robustez contra o sobreajuste. Todos os pesos lineares foram inicializados com a técnica de *Xavier Uniform Initialization* (GLOROT; BENGIO, 2010), enquanto os vieses foram inicializados em zero.

A saída da MLP corresponde a um vetor de 64 elementos que representa o *resíduo* a ser adicionado ao bloco central. Em vez de prever diretamente os coeficientes restaurados, o modelo utiliza uma conexão residual que combina a estimativa produzida pela rede com os coeficientes originais do bloco central, segundo:

$$\hat{y} = y_{\text{center}} + \lambda\, r(x), \qquad \lambda = 0.2, \tag{3.2}$$

em que y_{center} representa os 64 coeficientes do bloco central de entrada, $r(x)$ denota o vetor residual estimado pela rede, e λ é o fator de escala que regula a intensidade da correção. Esse mecanismo de *skip-connection*, inspirado nas *ResNets* (HE et al., 2016), favorece a preservação dos componentes de baixa frequência — usualmente mais estáveis sob compressão JPEG — e concentra o aprendizado nas componentes de alta frequência, que sofrem maior distorção durante o processo de quantização.

A mesma arquitetura foi treinada de forma independente para cada canal do espaço de cor YCbCr (Y, Cb e Cr), permitindo que cada rede se especializasse nas

características espectrais e estatísticas do canal correspondente.

3.7.2 Outras Arquiteturas Testadas

Durante o estudo, avaliaram-se variações arquiteturais com o objetivo de medir o impacto do contexto espacial e do tipo de predição. Testou-se uma MLP $64 \to 64$ sem contexto (predição direta dos coeficientes a partir do bloco degradado) e versões residuais sem vizinhança (1×1), além de funções de ativação alternativas, como *ReLU* e *SiLU*. Os resultados indicaram que a inclusão do contexto local (3×3) produziu ganhos consistentes em métricas como PSNR e SSIM. Entre as variantes, a `ContextResidualMLP` mostrou desempenho superior e maior estabilidade de treinamento, sendo adotada como configuração final.

3.7.3 Estratégias de Função de Perda

A escolha da função de perda exerceu papel central no comportamento de aprendizado. Consideraram-se três famílias principais: o erro quadrático médio (MSE), o erro absoluto médio (L1) e uma função de perda autoral, ponderada por frequência (*frequency-weighted MSE*), desenvolvida com o objetivo de enfatizar as perdas nos coeficientes de alta frequência da DCT.

O critério baseado na perda em frequência é dado por:

$$\mathcal{L}_{\text{freq}} = \frac{1}{64} \sum_{k=1}^{64} \tilde{w}_k \, (\hat{y}_k - y_k)^2, \qquad \tilde{w}_k = \frac{w_k}{\frac{1}{64} \sum_{j=1}^{64} w_j}, \qquad w_k = \text{linspace}(1, 100). \tag{3.3}$$

em que w_k representa o peso associado ao k-ésimo coeficiente DCT. Os pesos crescentes atribuem maior importância às altas frequências, correspondentes a detalhes finos da imagem, que são mais suscetíveis à perda durante a quantização JPEG.

Tal função de perda ponderada foi avaliada em combinação linear com a função de perda tradicional L1 da seguinte maneira:

$$\mathcal{L}_{\text{l1+freq}} = 0.7 \, \mathcal{L}_{\text{L1}} + 0.3 \, \mathcal{L}_{\text{freq}} \tag{3.4}$$

Na prática, a utilização da função L1, sem a combinação com $\mathcal{L}_{\text{freq}}$, apresentou o melhor desempenho geral, resultando em valores superiores de PSNR e MS-SSIM no conjunto de testes e reduzindo consideravelmente artefatos como *blocking* e *ringing* em cenários de maior compressão.

Os resultados quantitativos obtidos para cada função de perda avaliada podem ser consultados no Apêndice C.

3.8 PROCEDIMENTO DE TREINAMENTO

O treinamento foi conduzido de forma supervisionada sobre os *splits* previamente definidos. Cada partição (treinamento, validação e teste) foi carregada integralmente na memória RAM, com o objetivo de minimizar a latência da entrada e saída dos dados, bem como acelerar as iterações. O fornecimento de lotes foi realizado por `DataLoader` com `pin_memory` ativado, o que agilizou a transferência dos dados para a GPU.

Empregou-se o otimizador `AdamW`, com taxa de aprendizado inicial `lr`$= 5 \cdot 10^{-4}$ e `weight_decay`$= 10^{-4}$, valores que, mostraram bom equilíbrio entre rapidez de convergência e estabilidade numérica. O tamanho do lote foi configurado como 16.384, aproveitando de forma eficiente os recursos de memória disponíveis, e o treinamento foi executado por 30 épocas completas.

Ao final de cada época, avaliou-se o desempenho no conjunto de validação. O estado da rede associado à menor perda de validação foi salvo em disco, assegurando que o modelo utilizado nas etapas posteriores correspondia à melhor capacidade de generalização observada durante o processo.

O treinamento foi realizado de forma independente para cada canal Y, Cb e Cr, preservando arquitetura e hiperparâmetros. Tal estratégia permitiu que cada rede se especializasse nas propriedades espectrais e perceptuais do respectivo canal, contribuindo para o desempenho global do sistema no domínio da DCT.

4 RESULTADOS E DISCUSSÃO

Este capítulo apresenta e discute os resultados experimentais do método desenvolvido. A análise foi dividida em duas frentes complementares: (i) avaliação quantitativa, por meio de métricas objetivas de qualidade de imagem, e (ii) avaliação qualitativa, com exemplos visuais e estudo de casos. Sempre que pertinente, relatam-se medidas de variabilidade, intervalos de confiança e testes de hipótese, bem como análises por fator de qualidade JPEG (Q), por canal de cor (Y, Cb, Cr) e por sistema de cor (RGB).

4.1 PROTOCOLO DE AVALIAÇÃO

4.1.1 Métricas

Foram consideradas métricas com referência (PSNR e MS-SSIM) e métricas sem referência (NIQE e entropia). Para cada imagem i do conjunto de teste, calculou-se a métrica correspondente à imagem degradada $\left(m_i^{\text{in}}\right)$, à imagem aprimorada pelo modelo $\left(m_i^{\text{pred}}\right)$ e, quando aplicável, à imagem de referência de alta qualidade $\left(m_i^{\text{gt}}\right)$.

4.1.2 Estratificações

Os resultados são estratificados por fator de qualidade JPEG $Q \in \{10, 20, 30, 40, 50\}$ e por canal de cor (Y, Cb, Cr), quando pertinente, com o objetivo de evidenciar a robustez do método em diferentes regimes de quantização e em diferentes bandas espectrais.

4.1.3 Formulação dos Testes de Hipótese

Para avaliar se o método proposto proporciona melhorias estatisticamente significativas na qualidade das imagens, foram formulados testes de hipótese a partir dos ganhos por imagem em cada métrica de qualidade. Para cada fator de qualidade JPEG $Q \in \{10, 20, 30, 40, 50\}$ e para cada métrica considerada, define-se um ganho por imagem $d_{i,Q}$ entre a versão degradada (*input*) e a versão aprimorada pelo modelo (*pred*).

No caso das métricas em que valores mais altos indicam melhor qualidade (PSNR e MS-SSIM), o ganho é definido como

$$d_{i,Q} = m_{i,Q}^{\text{pred}} - m_{i,Q}^{\text{in}}.$$

Para o NIQE, em que valores menores são desejáveis, adotou-se

$$d_{i,Q} = m_{i,Q}^{\text{in}} - m_{i,Q}^{\text{pred}},$$

de modo que, em todas as métricas, valores positivos de $d_{i,Q}$ correspondem a melhoria em relação à imagem degradada.

Assim, para cada par (métrica, Q), obtém-se um conjunto de ganhos $d_{1,Q}, \ldots, d_{n,Q}$, com $n = 18$ imagens por fator de qualidade JPEG. Sobre esse conjunto, testa-se a hipótese nula de ausência de melhoria média,

$$H_0 : \mu_Q = 0,$$

contra a hipótese alternativa unilateral de que o método proporciona um ganho médio positivo,

$$H_1 : \mu_Q > 0,$$

em que μ_Q denota a média populacional dos ganhos para a métrica e fator de qualidade em questão.

Como o tamanho amostral n é inferior a 30 e o desvio-padrão populacional é desconhecido, aplica-se, para cada par (métrica, Q), o teste t de Student para uma amostra. O valor da estatística de teste é dado por

$$t = \frac{\bar{d}_Q}{s_Q/\sqrt{n}}, \tag{4.1}$$

em que $\bar{d}_Q$ é a média amostral dos ganhos, s_Q é o desvio-padrão amostral e $n = 18$ é o tamanho da amostra, resultando em $df = n - 1 = 17$ graus de liberdade.

Além do teste de significância, constroem-se intervalos de confiança de 95% para a média dos ganhos, a partir da distribuição t de Student com $df = 17$:

$$\text{IC}_{95\%} : \quad \bar{d}_Q \pm t_{0{,}975,17}\,\frac{s_Q}{\sqrt{n}}, \tag{4.2}$$

em que $t_{0{,}975,17}$ é o quantil da distribuição t correspondente à probabilidade acumulada de $97{,}5\%$.

Em todas as análises, adotou-se nível de significância $\mathrm{sig} = 0{,}05$. Valores de p unicaudal inferiores a sig foram interpretados como evidência de que o método proposto produz melhoria média estatisticamente significativa na métrica avaliada para o fator de qualidade considerado.

4.2 RESULTADOS QUANTITATIVOS

4.2.1 Resumo Global por Qualidade

As Tabelas 1 e 2 apresentam um resumo agregado das métricas empregadas na avaliação do modelo, estratificado pelo fator de qualidade JPEG.

Tabela 1 – Resumo estatístico dos resultados para as métricas full-reference (FR) por fator de qualidade JPEG: média e incerteza da média para os estágios degradado e aprimorado, no canal Y.

Q	MS-SSIM (in)		MS-SSIM (pred)		PSNR (in)		PSNR (pred)	
	média	inc	média	inc	média	inc	média	inc
10	0.943	0.003	0.961	0.003	28.21	0.72	29.47	0.80
20	0.974	0.002	0.981	0.002	30.89	0.85	32.20	0.93
30	0.983	0.001	0.987	0.001	31.38	0.66	32.71	0.72
40	0.986	0.001	0.989	0.0007	32.69	0.85	33.89	0.90
50	0.990	0.0005	0.992	0.0004	32.45	0.76	33.57	0.84

Tabela 2 – Resumo estatístico dos resultados para as métricas no-reference (NR) por fator de qualidade JPEG: média e incerteza padrão da média para os estágios referência (GT), degradado (in) e aprimorado (pred).

Q	Entropia (GT)		Entropia (in)		Entropia (pred)		NIQE (GT)		NIQE (in)		NIQE (pred)	
	média	inc	média	inc	média	inc	média	inc	média	inc	média	inc
10	7.26	0.13	6.49	0.22	7.06	0.16	3.40	0.33	6.36	0.40	4.73	0.34
20	7.34	0.11	7.09	0.15	7.27	0.12	3.09	0.21	4.59	0.22	4.08	0.23
30	7.39	0.11	7.29	0.14	7.34	0.13	3.06	0.32	3.87	0.27	3.75	0.27
40	7.29	0.08	7.25	0.09	7.28	0.08	3.02	0.29	3.44	0.24	3.57	0.30
50	7.58	0.04	7.56	0.04	7.53	0.04	2.77	0.22	3.02	0.22	3.30	0.26

As tabelas mostram que o modelo proposto foi capaz de melhorar consistentemente as métricas PSNR e MS-SSIM em todos os níveis de qualidade JPEG, com ganhos mais pronunciados em compressões mais severas (menores valores de Q). Em relação às métricas sem referência, observa-se uma redução significativa nos valores de NIQE para os níveis de qualidade mais baixos ($Q = 10$ e $Q = 20$), indicando uma melhoria perceptual nas imagens aprimoradas. Já a entropia das imagens aprimoradas aproxima-se da entropia das imagens de referência, sugerindo que o modelo é capaz de recuperar parte da informação perdida durante a compressão.

4.2.2 Análise estatística dos ganhos no canal Y

O canal de luminância (Y) é o mais relevante para a percepção visual humana. Logo, entende-se que a análise detalhada dos ganhos nesse canal seja crucial para avaliar o desempenho do modelo. A seguir, são apresentados os resultados estatísticos dos ganhos médios para as métricas PSNR, MS-SSIM e NIQE no canal Y, estratificados por fator de qualidade JPEG.

Tabela 3 – Ganho médio de PSNR por fator de qualidade JPEG no canal Y.

Q	n	Ganho médio (dB)	IC 95%	t	p (unicaudal)
10	18	1.253	1,002 – 1,504	10,55	3,5e-09
20	18	1.311	1,081 – 1,540	12,05	4,7e-10
30	18	1.322	1,124 – 1,520	14,08	4,2e-11
40	18	1.198	1,030 – 1,365	15,08	1,4e-11
50	18	1.121	0,925 – 1,317	12,09	4,5e-10

A Tabela 3 apresenta o ganho médio da métrica PSNR no canal de luminância para cada fator de qualidade JPEG. Analisando-a, constata-se um ganho substancial na relação sinal-ruído das imagens aprimoradas em relação às imagens comprimidas, com todos os ganhos médios situando-se de $1{,}1$ a $1{,}3\,\mathrm{dB}$, bem como os intervalos de confiança de 95% acima de zero. Corroborando o bom desempenho do modelo para essa métrica, observam-se valores de t elevados e valores-p unidimensionais muito pequenos ($p < 10^{-9}$), o que nos permite rejeitar a hipótese nula. Entretanto, embora o ganho se mantenha em torno de $1\,\mathrm{dB}$ mesmo para compressões mais leves ($Q = 50$), observa-se uma leve tendência de redução do ganho à medida que a qualidade JPEG aumenta, o que é compatível com a ideia de que imagens menos degradadas oferecem uma menor margem para melhorias no domínio espectral.

Tabela 4 – Ganho médio de MS-SSIM por fator de qualidade JPEG no canal Y.

Q	n	Ganho médio	IC 95%	t	p (unicaudal)
10	18	0,018	0,014 – 0,023	8,14	1,4e-07
20	18	0,007	0,006 – 0,009	9,92	8,7e-09
30	18	0,004	0,003 – 0,005	10,19	5,9e-09
40	18	0,003	0,002 – 0,004	10,75	2,7e-09
50	18	0,002	0,002 – 0,003	12,65	2,2e-10

Na Tabela 4 são apresentados os resultados da métrica MS-SSIM. Nota-se que a rede neural apresenta ganhos médios positivos para todos os valores de Q, com todos os intervalos de confiança de 95% acima de zero, bem como $p \ll 0{,}05$ para todos os níveis de qualidade. Para a compressão mais severa ($Q = 10$), o ganho médio é próximo de $0{,}018$, decrescendo gradualmente até cerca de $0{,}002$ para $Q = 50$. Embora os ganhos médios da MS-SSIM sejam inferiores aos ganhos da PSNR — salvo as diferenças de proporções entre as métricas — tais variações indicam uma melhoria estatisticamente significativa na similaridade estrutural entre as imagens restauradas e suas referências, em comparação com as imagens degradadas; sobretudo em cenários de alta compressão. Em termos qualitativos, os resultados sugerem que o modelo é capaz de restaurar detalhes estruturais perdidos no processo de compressão, ainda que o ganho se torne menos perceptível quando o nível de deterioração da imagem é menor.

Tabela 5 – Redução média de NIQE por fator de qualidade JPEG no canal Y.

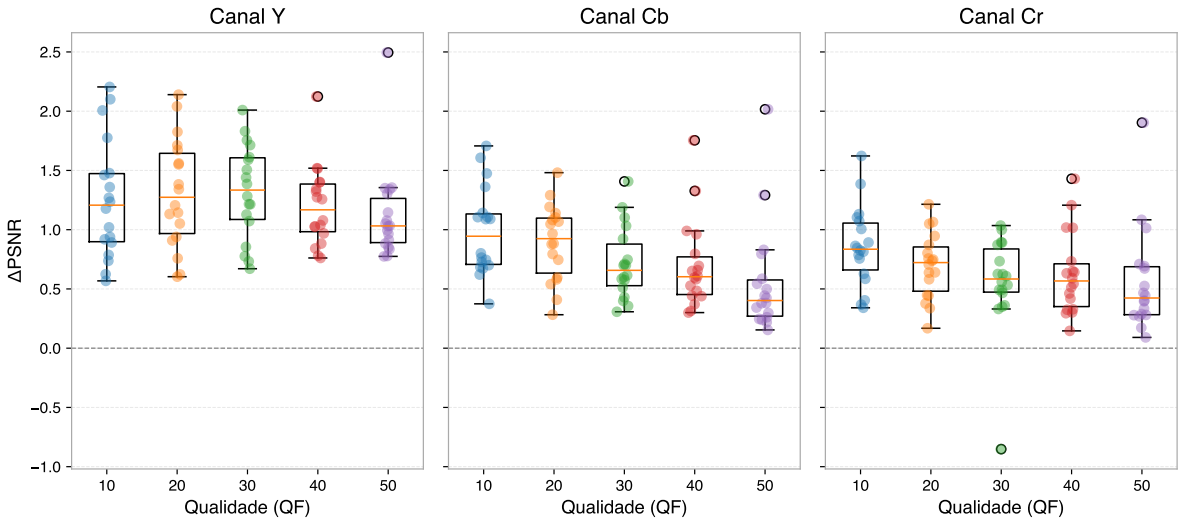
Q	n	Ganho médio	IC 95% (ganho)	t	p (unicaudal)
10	18	1,630	1,396 – 1,863	14,73	2,1e-11
20	18	0,501	0,300 – 0,702	5,27	3,1e-05
30	18	0,123	-0,026 – 0,272	1,74	5,0e-02
40	18	-0,134	-0,318 – 0,050	-1,54	9,3e-01
50	18	-0,286	-0,434 – -0,139	-4,11	1,0e+00

A Tabela 5 resume os resultados para a métrica NIQE. Devido às propriedades da métrica, um ganho positivo nessa tabela corresponde a uma redução média de NIQE (melhoria) e um ganho negativo indica piora da qualidade da imagem. Para $Q = 10$ e $Q = 20$, observam-se ganhos médios expressivos ($1{,}63$ e $0{,}50$, respectivamente), com intervalos de confiança de 95% estritamente positivos e valores-p muito pequenos, o que evidencia uma melhoria perceptual clara em relação às imagens comprimidas. Para $Q = 30$, o ganho médio ainda é positivo ($0{,}123$); no entanto, o intervalo de confiança inclui zero, indicando que não há evidência estatística suficiente para concluir que a hipótese nula é falsa. Já para $Q = 40$ e $Q = 50$, os ganhos médios tornam-se negativos. No caso de $Q = 40$, o intervalo de confiança ainda inclui zero, sugerindo que não há evidência estatística forte nem de melhoria nem de piora. Já para $Q = 50$, todavia, o intervalo de confiança encontra-se inteiramente abaixo de zero, o que indica que o modelo tende a comprometer a naturalidade das imagens.

4.2.3 Distribuição dos ganhos por canal e qualidade

4.2.3.1 PSNR

Figura 17 – Distribuição dos ganhos de PSNR por canal (Y, Cb, Cr) e por fator de qualidade JPEG.



Fonte: De autoria própria.

Os resultados de ganho de ΔPSNR apresentados na Figura 17 mostram um comportamento coerente com a teoria da DCT e com o processo de quantização do padrão JPEG.

Em qualidades mais baixas ($QF = 10, 20$), as matrizes de quantização JPEG aplicam fatores elevados, resultando em perdas agressivas principalmente nas bandas de alta frequência. Como o PSNR mede essencialmente a relação sinal-ruído (na

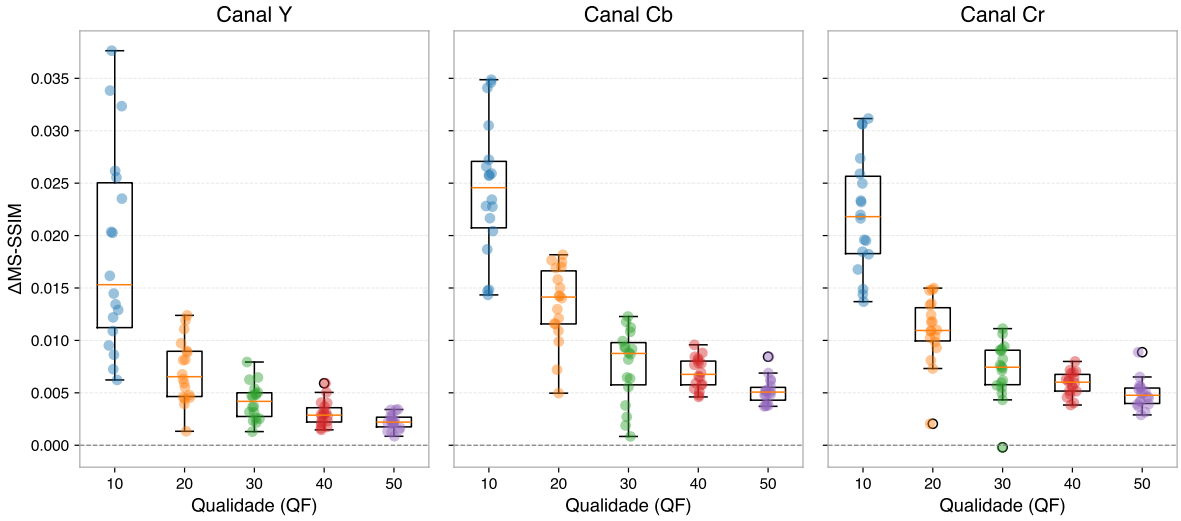
prática, a intensidade do erro de reconstrução no domínio espacial), a correção desses coeficientes degradados leva a reduções significativas do erro quadrático médio, o que se traduz nos maiores ganhos observados de PSNR nessas qualidades.

À medida que o fator de qualidade aumenta ($QF \geq 30$), a quantização torna-se gradativamente menos severa e o erro introduzido pelo JPEG é mais discreto. Consequentemente, o espaço para correção efetiva diminui, refletindo-se em ganhos menores e mais homogêneos entre as imagens.

De modo geral, o comportamento decrescente dos ganhos de PSNR com o aumento de QF é compatível com o modelo de degradação JPEG e evidencia que o modelo aprende, de forma eficaz, a corrigir distorções introduzidas pela quantização, sobretudo aquelas que afetam componentes de alta frequência.

4.2.3.2 MS-SSIM

Figura 18 – Distribuição dos ganhos de MS-SSIM por canal (Y, Cb, Cr) e por fator de qualidade JPEG.



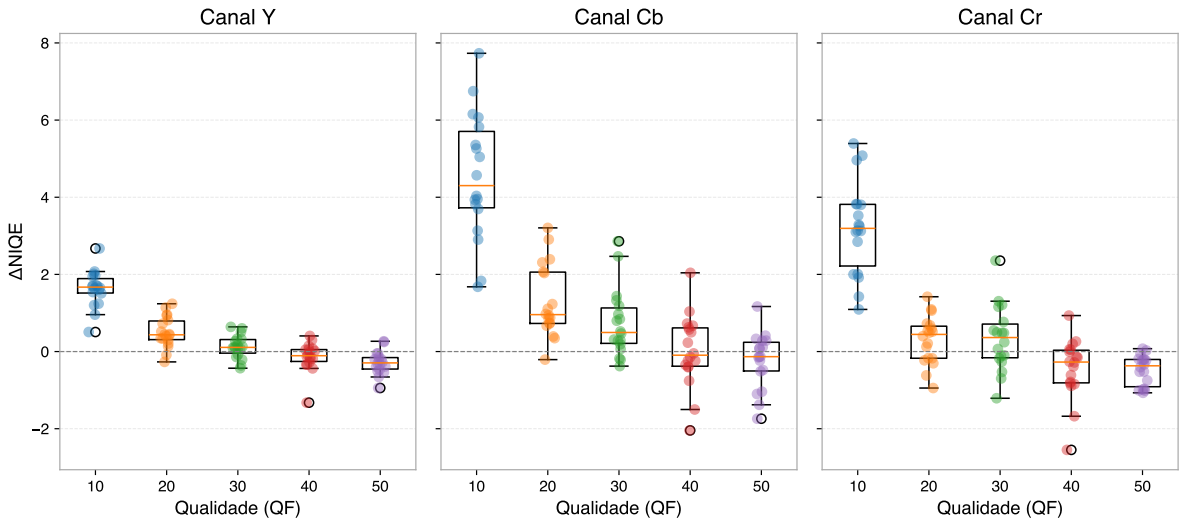
Fonte: De autoria própria.

Ao analisar a distribuição dos ganhos de ΔMS-SSIM presentes na Figura 18, observa-se, para o canal Y, que os ganhos são mais pronunciados para compressões mais agressivas ($Q = 10$). Esse comportamento é esperado, uma vez que a maior parte da informação estrutural (bordas, texturas e luminância) encontra-se nesse canal, e a compressão JPEG degrada tais componentes de forma mais perceptível. À medida que a qualidade aumenta ($Q \geq 20$), a magnitude dos ganhos se reduz, embora permaneça sempre positiva, indicando que o método continua sendo capaz de aprimorar a qualidade estrutural das imagens, mesmo em cenários de compressão menos severa.

Nos canais Cb e Cr, as melhorias seguem o mesmo padrão qualitativo, porém com magnitudes ligeiramente maiores do que no canal Y para todas as qualidades. Esse fenômeno pode ser atribuído ao fato de que os canais de crominância possuem coeficientes DCT mais esparsos, tornando-os mais suscetíveis a melhorias quando o modelo é aplicado. Assim, o modelo consegue recuperar detalhes cromáticos que foram significativamente afetados pela quantização, resultando em ganhos médios mais elevados nesses canais.

4.2.3.3 NIQE

Figura 19 – Distribuição dos ganhos de NIQE por canal (Y, Cb, Cr) e por fator de qualidade JPEG.



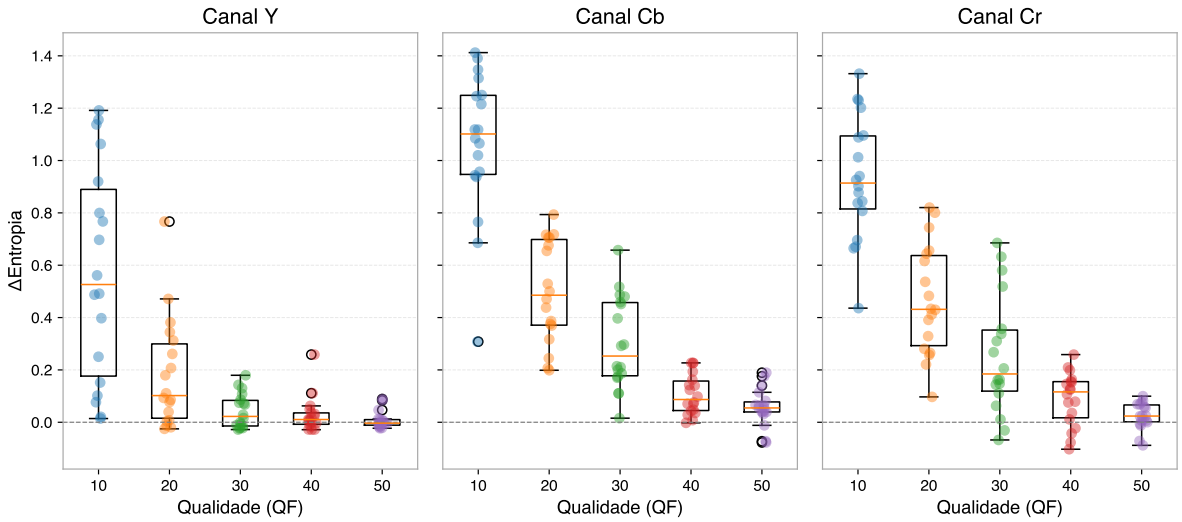
Fonte: De autoria própria.

Os resultados de ΔNIQE na Figura 19 exibem melhorias claras apenas em condições de compressão intensa ($QF = 10, 20$), diminuindo progressivamente conforme a quantização JPEG se torna menos severa. Esse comportamento é consistente com a natureza do NIQE, que responde fortemente a artefatos de alta frequência, como blocos, *ringing* e perda de textura, todos diretamente relacionados à quantização dos coeficientes da DCT.

A arquitetura `ContextResidualMLP`, ao operar no domínio da DCT com vizinhança 3×3, é particularmente eficaz em corrigir artefatos introduzidos pela quantização agressiva, como blocagem. Assim, quando a degradação é substancial, a rede reduz de maneira significativa as irregularidades estatísticas detectadas pelo NIQE. Entretanto, para qualidades mais altas ($QF \geq 30$), a perturbação nos coeficientes DCT é menor e o NIQE torna-se menos sensível às melhorias proporcionadas pelo modelo, resultando em ganhos médios próximos de zero ou até negativos.

4.2.3.4 Entropia

Figura 20 – Distribuição dos ganhos de entropia por canal (Y, Cb, Cr) e por fator de qualidade JPEG.



Fonte: De autoria própria.

Os resultados de ΔEntropia mostram que o modelo aumenta a incerteza estatística diminuída pela quantização JPEG, especialmente em baixos níveis de qualidade ($QF = 10, 20$). O que nos leva a concluir que a imagem aprimorada possui maior quantidade de informação, aproximando-se mais da referência original.

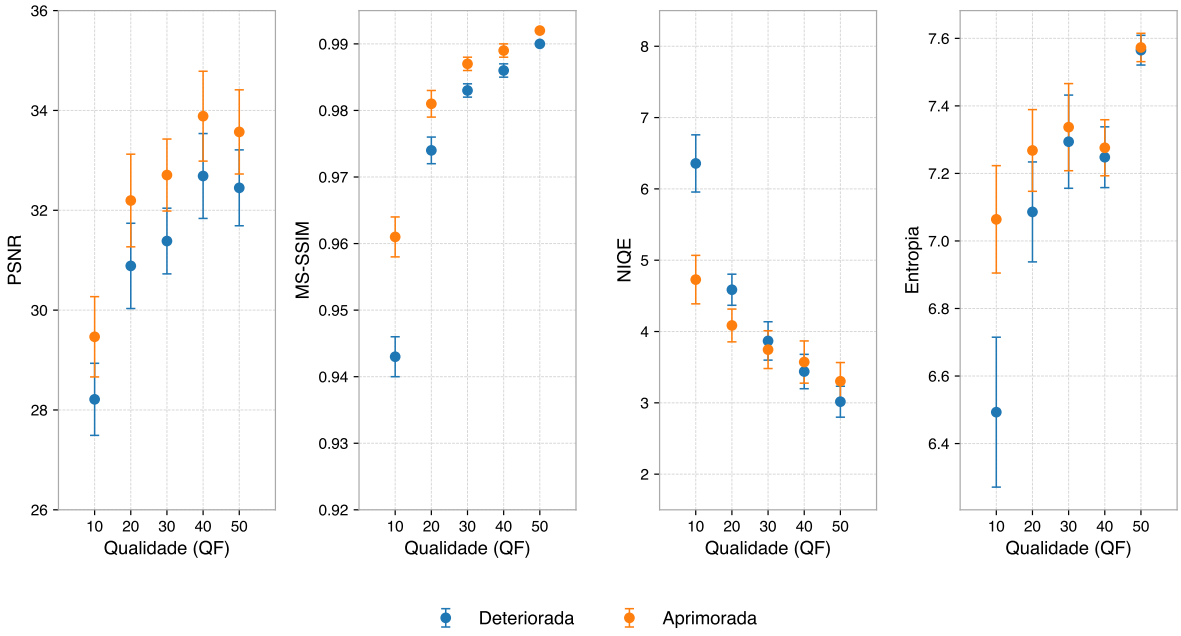
Nessa faixa, a compressão remove componentes de alta frequência de forma desigual, diminuindo a desordem local dos coeficientes, efeito que a `ContextResidualMLP` consegue mitigar ao restaurar energia em padrões estruturados da vizinhança 3×3. Para qualidades mais altas ($QF \geq 30$), a entropia dos coeficientes degradados já é próxima da referência, e o modelo produz apenas ajustes sutis, resultando em ganhos pequenos ou próximos de zero, consistentes com a menor degradação.

4.2.4 Comparativo Geral das Métricas por Qualidade

A seguir, a Figura 21 apresenta um comparativo geral das métricas de qualidade entre as imagens de entrada (deteriorada) e as imagens preditas (aprimoradas) por fator de qualidade JPEG, consolidando os resultados quantitativos discutidos anteriormente.

O conjunto de gráficos exposto na Figura 21 evidencia que o modelo desenvolvido, `ContextResidualMLP`, produz melhorias consistentes e estatisticamente robustas nas métricas de qualidade, sobretudo em regimes de compressão JPEG mais severa. Para o PSNR, observa-se ganho médio de aproximadamente 1 dB em todas as qualidades, indicando redução efetiva do erro quadrático entre a imagem aprimorada e a referência. O comportamento do MS-SSIM reforça essa tendência: as curvas da versão

Figura 21 – Comparativo geral das métricas de qualidade entre as imagens de entrada (deteriorada) e as imagens preditas (aprimoradas) por fator de qualidade JPEG, para o canal de luminância.



Fonte: De autoria própria.

aprimorada aproximam-se dos valores de referência à medida que a rede restaura estruturas locais suprimidas pela quantização DCT.

O NIQE, sensível a irregularidades estatísticas e artefatos de alta frequência, apresenta reduções proeminentes em $QF = 10, 20$, coerentes com a maior intensidade de quantização nesses cenários. Por fim, a entropia revela que o modelo recupera informação perdida nos coeficientes DCT, especialmente no canal Y, sem induzir ruído adicional em altas qualidades. No conjunto, os resultados confirmam que a aprendizagem residual no domínio da DCT é eficaz para reconstrução de detalhes locais e supressão de artefatos de compressão.

4.3 RESULTADOS QUALITATIVOS

Nesta seção, selecionaram-se, para cada valor de Q, exemplos com características visuais distintas, incluindo texturas finas, arestas nítidas e regiões suaves. Para cada caso, apresentam-se tripletos contendo a imagem de referência, a imagem degradada e a imagem aprimorada, além de recortes de 128×128 em regiões de interesse. Discute-se o efeito dos modelos sobre a imagem deteriorada, incluindo o aprimoramento geral e a mitigação de artefatos típicos, como blocagem, *ringing* e *mosquito noise*.

4.3.1 Compressão com perdas severa

Como discutido anteriormente, as redes neurais desenvolvidas demonstraram um desempenho expressivo na restauração de imagens comprimidas com altos níveis de degradação. As figuras a seguir apresentam a predição obtida para as imagens do conjunto de testes com $Q = 10$. Cada canal de cor foi aprimorado por meio da versão correspondente do modelo `ContextResidualMLP`, treinado exclusivamente com os dados desse canal. As imagens preditas correspondem à reconstrução final, combinando os três canais aprimorados (Y, Cb e Cr).

Figura 22 – Comparativo geral entre a imagem original, degradada e predita do item 1 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 10$.

(a) Imagem original (b) Imagem degradada (c) Imagem predita

Fonte: De autoria própria.

Figura 23 – Comparativo da região de interesse do item 1 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 10$.

(a) Imagem degradada
$\mathrm{PSNR} = 28{,}27$; $\mathrm{MS-SSIM} = 0{,}9204$;
$\mathrm{NIQE} = 6{,}435$; $\mathrm{Entropia} = 6{,}639$.

(b) Imagem predita
$\mathrm{PSNR} = 29{,}21$; $\mathrm{MS-SSIM} = 0{,}9439$;
$\mathrm{NIQE} = 4{,}463$; $\mathrm{Entropia} = 7{,}336$.

Fonte: De autoria própria.

Analisando as Figuras 22 e 23, nota-se a efetividade dos modelos desenvolvidos na mitigação dos artefatos resultantes da compressão. A imagem predita apresenta uma redução significativa no efeito de blocagem, com transições tonais mais suaves entre os blocos DCT adjacentes, indicando que os modelos inferiram corretamente as correlações entre os blocos DCT vizinhos. Ao redor das bordas do olho e dos

cílios, a versão degradada apresenta halos suaves e oscilações, características do artefato *ringing*. Na versão aprimorada, esses halos foram substancialmente reduzidos, com contornos mais limpos e transições mais estáveis. Isso evidencia que o modelo conseguiu recompor altas frequências de forma coerente com a estrutura local. Conclui-se que a imagem predita preservou a naturalidade fotográfica, sem introduzir padrões artificiais ou texturas sintéticas irreais. Não há evidências visíveis de nitidez extrema ou halos secundários, o que demonstra um bom equilíbrio entre suavização e restauração.

Figura 24 – Comparativo geral entre a imagem original, degradada e predita do item 9 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 10$.

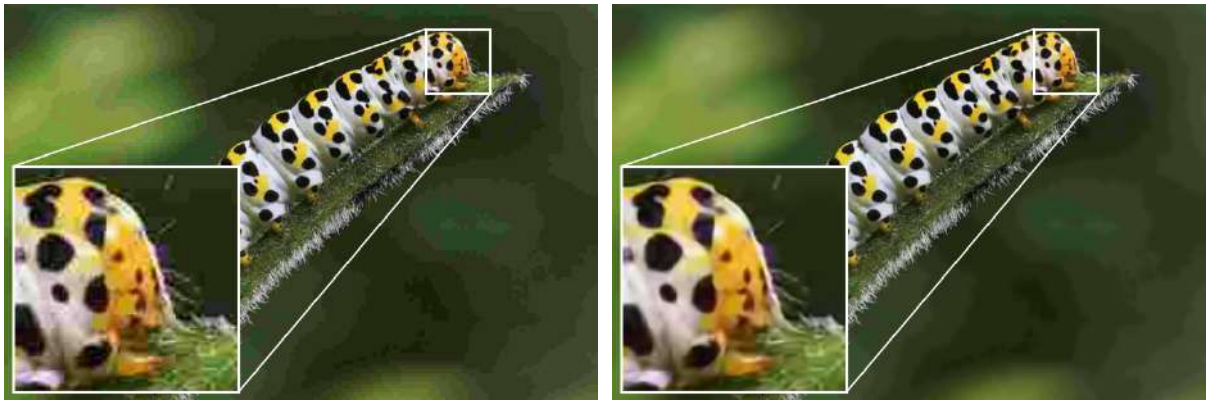
(a) Imagem original (b) Imagem degradada (c) Imagem predita

Fonte: De autoria própria.

Figura 25 – Comparativo da região de interesse do item 9 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 10$.

(a) Imagem degradada.
$\mathrm{PSNR} = 29{,}17$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9348$;
$\mathrm{NIQE} = 6{,}353$; $\mathrm{Entropia} = 5{,}438$.

(b) Imagem predita
$\mathrm{PSNR} = 30{,}44$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9671$;
$\mathrm{NIQE} = 4{,}278$; $\mathrm{Entropia} = 6{,}629$.



Fonte: De autoria própria.

Ao analisar os resultados do aprimoramento da imagem de número 9 do conjunto de testes, ilustrados nas Figuras 24 e 25, observa-se novamente um resultado positivo dos modelos de melhoria. Os modelos mitigaram de maneira expressiva o efeito de *mosquito noise*, resultando em uma textura da pele da lagarta com contraste melhor preservado. Contudo, devido à janela de 9 blocos de coeficientes DCT utilizada no treinamento, isto é, uma região de 192×192 píxeis, ainda se observam efeitos da

compressão em áreas extensas de baixa frequência, como no plano de fundo da imagem. Ainda assim, a restauração gera ganhos sem comprometer o realismo visual: não há excesso de nitidez artificial, as cores permanecem consistentes com a referência degradada e os detalhes recuperados não apresentam indícios evidentes de alucinação estrutural, confirmando a coerência estrutural com a imagem original.

4.3.2 Compressão com perdas moderada

A seguir, apresentam-se os resultados do aprimoramento de imagens com $Q = 30$, 40 e 50. As imagens preditas correspondem à reconstrução final obtida pela combinação dos três canais (Y, Cb e Cr) em uma imagem RGB, conforme a Equação 2.2.

Figura 26 – Comparativo geral entre a imagem original, degradada e predita do item 49 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 30$.

(a) Imagem original (b) Imagem degradada (c) Imagem predita

Fonte: De autoria própria.

Figura 27 – Comparativo da região de interesse do item 49 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 30$.

(a) Imagem degradada.
$\mathrm{PSNR} = 33{,}01$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9797$;
$\mathrm{NIQE} = 3{,}747$; $\mathrm{Entropia} = 7{,}301$.

(b) Imagem predita
$\mathrm{PSNR} = 34{,}23$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9862$;
$\mathrm{NIQE} = 3{,}653$; $\mathrm{Entropia} = 7{,}409$.

Fonte: De autoria própria.

Figura 28 – Comparativo geral entre a imagem original, degradada e predita do item 70 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 40$.

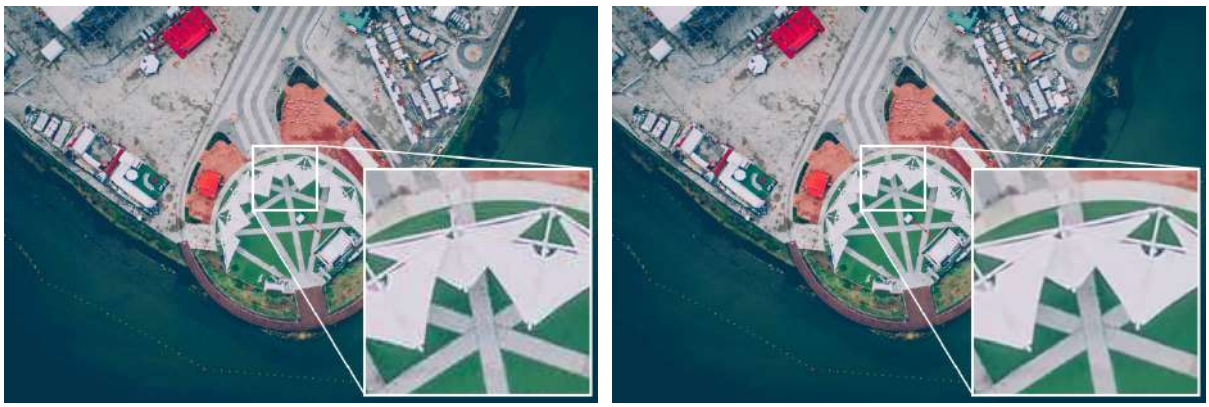
(a) Imagem original (b) Imagem degradada (c) Imagem predita

Fonte: De autoria própria.

Figura 29 – Comparativo da região de interesse do item 70 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 40$.

(a) Imagem degradada.
$\mathrm{PSNR} = 31{,}00$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9795$;
$\mathrm{NIQE} = 2{,}845$; $\mathrm{Entropia} = 6{,}891$.

(b) Imagem predita
$\mathrm{PSNR} = 32{,}32$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9836$;
$\mathrm{NIQE} = 2{,}851$; $\mathrm{Entropia} = 6{,}953$.



Fonte: De autoria própria.

Figura 30 – Comparativo geral entre a imagem original, degradada e predita do item 89 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 50$.

(a) Imagem original (b) Imagem degradada (c) Imagem predita

Fonte: De autoria própria.

Assim como nos resultados quantitativos, observa-se que o desempenho dos modelos propostos em imagens com perdas moderadas não foi tão expressivo quanto nas imagens submetidas a compressão severa, conforme ilustrado nas Figuras 27, 29 e 31.

Figura 31 – Comparativo da região de interesse do item 89 do conjunto de testes, comprimido pelo padrão JPEG com $Q = 50$.

(a) Imagem degradada.
$\mathrm{PSNR} = 40{,}23$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9904$;
$\mathrm{NIQE} = 5{,}510$; $\mathrm{Entropia} = 7{,}498$.

(b) Imagem predita
$\mathrm{PSNR} = 42{,}72$; $\mathrm{MS} - \mathrm{SSIM} = 0{,}9932$;
$\mathrm{NIQE} = 6{,}169$; $\mathrm{Entropia} = 7{,}494$.

Fonte: De autoria própria.

Nota-se uma redução consistente de artefatos característicos da compressão JPEG, especialmente dos artefatos de alta frequência, *mosquito noise* e *ringing*. Entretanto, a restauração de detalhes finos nessas faixas de qualidade apresenta limitações. Embora os modelos consigam suavizar irregularidades e estabilizar bordas, não se observa um ganho significativo na nitidez global, tampouco uma recuperação profunda das microestruturas de padrões geométricos presentes na imagem original.

Tal comportamento é coerente com a natureza local da arquitetura baseada em blocos DCT e vizinhanças 3×3, e indica que os modelos desenvolvidos apresentam maior aptidão para mitigar artefatos introduzidos por compressões agressivas, em que as perdas estruturais são mais evidentes, do que para reconstruir sutilezas de alto nível em imagens de qualidade mediana, nas quais a degradação é mais sutil e, consequentemente, mais difícil de modelar e corrigir.

4.4 EFICIÊNCIA COMPUTACIONAL

A eficiência computacional do modelo `ContextResidualMLP` para o canal Y foi avaliada por meio da medição do tempo médio de inferência para imagens com resolução de 1200×800 píxeis em uma placa de vídeo NVIDIA Tesla A100, disponibilizada no ambiente Google Colab Pro. A Tabela 6 apresenta os resultados obtidos para diferentes fatores de qualidade JPEG.

Observa-se que o tempo de inferência é similar em todos os fator de qualidade JPEG avaliados, com média geral de 0,03434 segundos por imagem. Esse valor corresponde a um *throughput* de cerca de 29 imagens por segundo, relativo ao ambiente mencionado. Embora a placa de vídeo utilizada seja de alto desempenho, o modelo

Tabela 6 – Tempo médio de inferência por imagem de 1200×800 píxeis, por fator de qualidade JPEG, para o canal de luminância.

Qualidade	Média de duração da inferência (s)	Desv. pad. (s)	Inc. da média (s)
10	0.0343	0.0003	0.0001
20	0.0343	0.0005	0.0001
30	0.0342	0.0005	0.0001
40	0.0345	0.0005	0.0001
50	0.0344	0.0005	0.0001

é relativamente leve, o que sugere que tempos de inferência da mesma ordem de grandeza podem ser obtidos também em hardwares de porte intermediário. O tamanho do modelo completo em disco é de aproximadamente 25 MB, contendo cerca de 2.203.200 parâmetros treináveis. Tais valores refletem que a arquitetura desenvolvida equilibra de forma adequada a complexidade do modelo com sua capacidade de generalização e desempenho na tarefa de aprimoramento de imagens comprimidas.

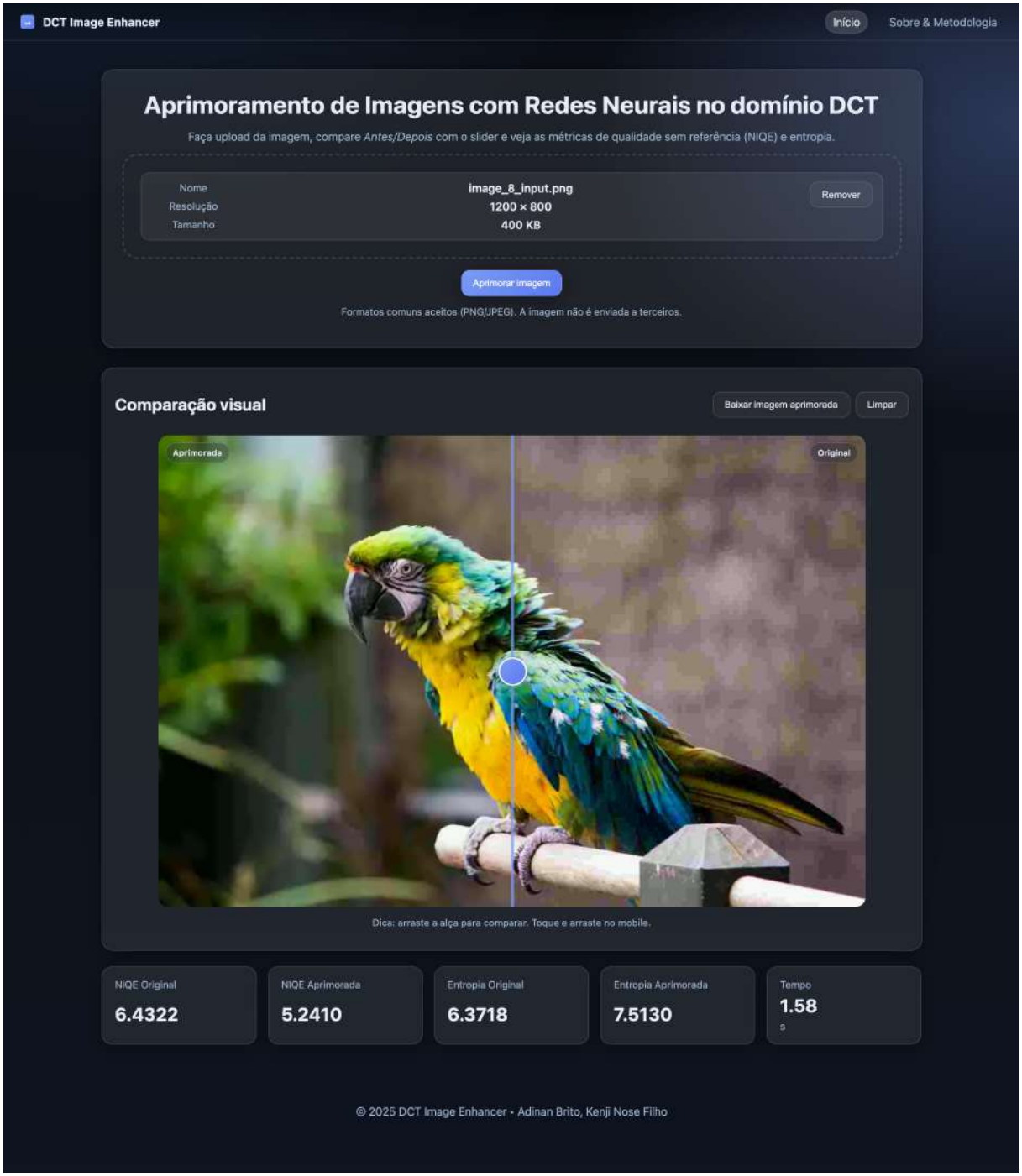
4.5 INTERFACE WEB PARA DEMONSTRAÇÃO DO APRIMORAMENTO

Paralelamente ao desenvolvimento dos modelos de aprimoramento de imagens JPEG, foi criada uma interface web interativa que permite aos usuários testar o modelo `ContextResidualMLP` em suas próprias imagens comprimidas. Capturas de tela da interface desenvolvida são exibidas nas Figuras 32 e 33.

A solução utiliza o framework FastAPI (TIANGOLO, 2025) para Python no servidor e HTML/CSS/JavaScript no cliente, proporcionando uma experiência de uso facilitada e responsiva. Ela possibilita o envio de imagens nos formatos JPEG e PNG, a visualização dos resultados aprimorados em poucos segundos, o download das imagens processadas e a comparação lado a lado entre a versão original e a aprimorada. Adicionalmente, a interface exibe as métricas de qualidade NIQE e entropia calculadas para o canal de luminância de cada imagem, proporcionando uma mensuração quantitativa do aprimoramento realizado.

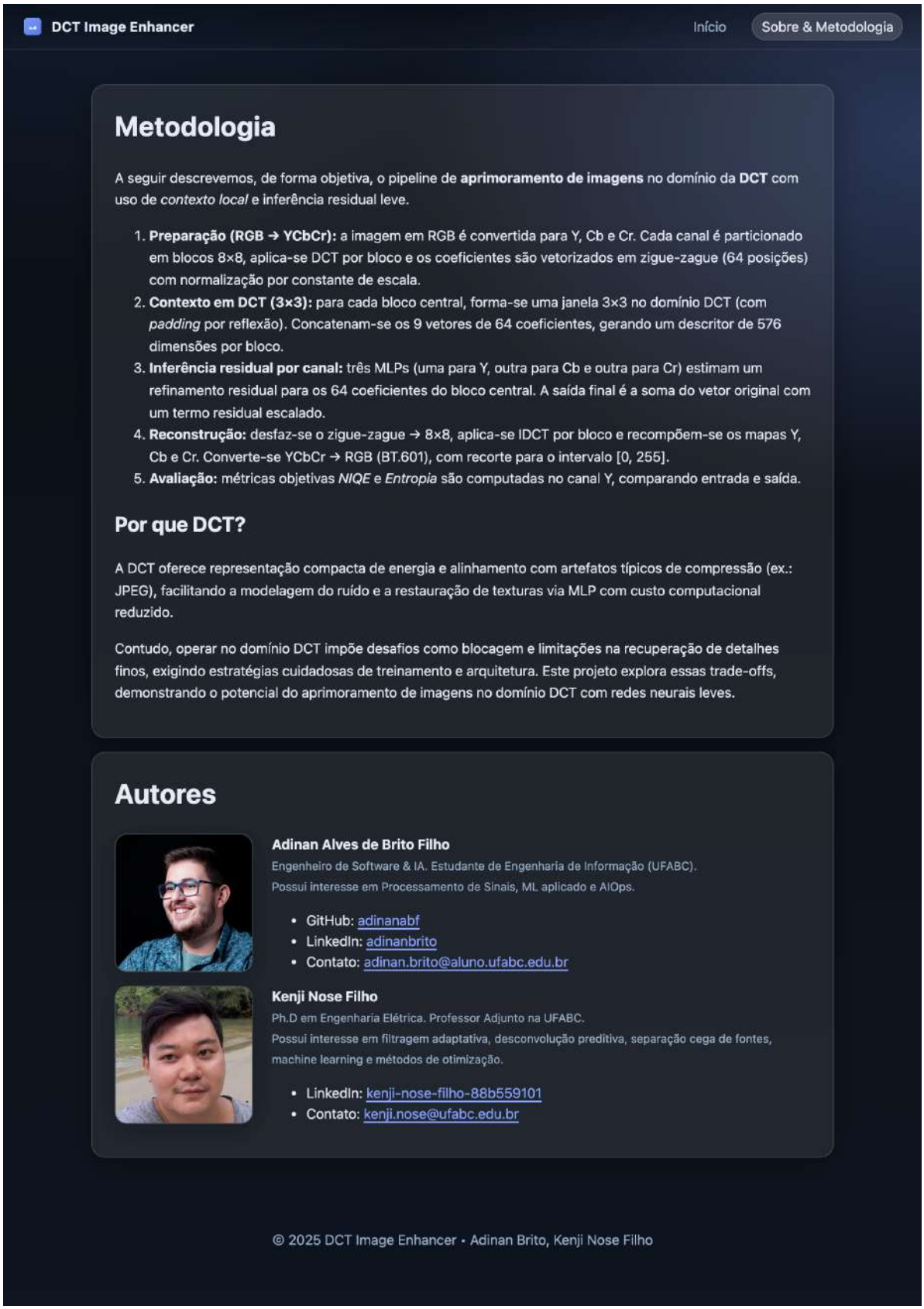
A ferramenta tem como objetivo facilitar a disseminação e o uso prático dos modelos desenvolvidos, permitindo que usuários de diferentes níveis técnicos experimentem, de forma acessível, os benefícios dos modelos de aprimoramento de imagens elaborados neste trabalho. O código-fonte completo da interface, bem como os arquivos dos modelos desenvolvidos, estão disponíveis em: github.com/adinanabf/dct-image-enhancer.

Figura 32 – Página inicial da interface web para demonstração do aprimoramento de imagens JPEG utilizando os modelos `ContextResidualMLP`.



Fonte: De autoria própria.

Figura 33 – Página "Sobre e Metodologia"da interface web para demonstração do aprimoramento de imagens JPEG utilizando os modelos `ContextResidualMLP`.



Fonte: De autoria própria.

5 CONCLUSÃO

Neste trabalho, foi proposta uma rede neural artificial, denominada `ContextResidualMLP`, com o objetivo de aprimorar imagens comprimidas no formato JPEG, atuando diretamente no domínio da DCT. A abordagem adotada consistiu em treinar o modelo para estimar o resíduo de correção a ser somado aos coeficientes DCT degradados de imagens comprimidas, utilizando uma arquitetura do tipo MLP que incorpora informações contextuais dos blocos vizinhos para melhorar a precisão das predições. A metodologia envolveu a criação de um conjunto de dados diversificado, abrangendo diferentes níveis de qualidade JPEG, e a implementação de técnicas de pré-processamento e normalização dos dados para otimizar o treinamento do modelo. A avaliação do desempenho do `ContextResidualMLP` foi realizada por meio de métricas quantitativas, como PSNR e MS-SSIM, além de análises qualitativas das imagens aprimoradas.

Como demonstrado no Capítulo 4, o aprimoramento de imagens no domínio da DCT mostrou-se uma estratégia eficaz para mitigar os artefatos introduzidos pela compressão com perdas, conforme as métricas consideradas e o teste t de Student adotado neste trabalho. Os modelos desenvolvidos foram capazes de aprender padrões estatísticos de degradação dos coeficientes DCT introduzidos pela compressão JPEG, atenuando os efeitos perceptuais da perda de informação, como blocagem e *mosquito noise*, sobretudo em imagens submetidas a altas taxas de compressão. Além disso, a baixa quantidade de parâmetros dos modelos – em comparação a modelos convolucionais profundos empregados no aprimoramento de imagens no domínio espacial – e a eficiência da etapa de predição, avaliada em ambiente com aceleração por GPU, indicam que a solução proposta apresenta baixo custo computacional relativo, considerando o escopo experimental deste trabalho, o que favorece sua disseminação e potencial uso em cenários aplicados que demandem inferência rápida.

Em termos de contribuições acadêmicas, este trabalho agrega valor ao campo de estudo do processamento de imagens e compressão com perdas ao explorar uma abordagem menos frequente na literatura, predominantemente dominada por métodos convolucionais no domínio espacial. A proposta do `ContextResidualMLP` amplia o leque de abordagens possíveis para o problema de aprimoramento de imagens comprimidas, oferecendo uma alternativa eficaz e eficiente em relação aos métodos tradicionais baseados em filtros ou técnicas mais complexas de *Deep Learning*, como arquiteturas convolucionais profundas ou modelos baseados em atenção, sob a perspectiva do custo computacional.

Como continuidade deste trabalho, seria interessante investigar arquiteturas neurais estruturalmente mais sofisticadas, como Redes Neurais Convolucionais (CNN)

e *Transformers*, bem como novas metodologias de treinamento em que cada fator de qualidade JPEG disponha de uma rede específica de aprimoramento, especializando o modelo na mitigação dos artefatos característicos de cada nível de compressão. Também se mostra promissora a análise da relação entre a classe da imagem (por exemplo, paisagens, retratos, padrões geométricos ou tipografia) e os efeitos da compressão DCT com perdas em cada uma dessas categorias, abrindo espaço para o desenvolvimento de modelos especializados por tipo de conteúdo.

A interface web desenvolvida para a demonstração do aprimoramento agrega valor didático e tecnológico a este trabalho, materializando os conhecimentos de programação de sistemas adquiridos ao longo do Bacharelado em Engenharia de Informação em uma solução concreta. Essa interface não apenas torna o estudo mais acessível e compreensível para um público amplo, como também aproxima o conhecimento científico da população, transformando um trabalho acadêmico em um produto que pode, potencialmente, beneficiar a sociedade.

Por fim, a realização deste Trabalho de Graduação representa a síntese da trajetória acadêmica construída ao longo do Bacharelado em Engenharia de Informação. A solução proposta é, ao mesmo tempo, simples e complexa, refletindo a riqueza de detalhes de uma jornada de aprendizado vivida em conjunto com a comunidade acadêmica da Universidade Federal do ABC: uma experiência cuja essência nem mesmo uma compressão com perdas no domínio da DCT seria capaz de deteriorar.

REFERÊNCIAS

AGARAP, A. F. Deep Learning using Rectified Linear Units (ReLU). *arXiv e-prints*, p. arXiv:1803.08375, mar. 2018. Citado na página 17.

AGUSTSSON, E.; TIMOFTE, R. Ntire 2017 challenge on single image super-resolution: Dataset and study. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. [S.l.: s.n.], 2017. Citado na página 23.

AHMED, N.; NATARAJAN, T.; RAO, K. Discrete cosine transform. *IEEE Transactions on Computers*, C-23, n. 1, p. 90–93, 1974. Disponível em: <https://ieeexplore.ieee.org/document/1672377>. Citado na página 2.

ALGOLITH. *Before and After Mosquito Noise*. [S.l.]: Ziff Davis, 2025. Digital image on PCMag. Image illustrating mosquito noise reduction, originally published in "Definition of mosquito noise"on PCMag.com. URL: https://www.pcmag.com/encyclopedia/term/mosquito-noise. Citado na página 15.

ASSI, A.; SHANWAR, B.; ZARKA, N. *Detection of Abandoned Objects in Crowded Environments*. [S.l.], 2016. Disponível em: <https://www.researchgate.net/publication/304625497_Detection_of_Abandoned_Objects_in_Crowded_Environments>. Citado na página 7.

AYDIN, O.; CİNBİŞ, R. Single-image super-resolution analysis in dct spectral domain. *Balkan Journal of Electrical and Computer Engineering*, v. 8, p. 209–217, 07 2020. Disponível em: <https://www.researchgate.net/publication/343320619_Single-Image_Super-Resolution_Analysis_in_DCT_Spectral_Domain>. Citado na página 1.

BENYAMINOVICH, S.; HADAR, O.; KAMINSKY, E. Optimal transrating via dct coefficients modification and dropping. In: *ITRE 2005. 3rd International Conference on Information Technology: Research and Education, 2005.* [S.l.: s.n.], 2005. p. 100–104. Citado na página 11.

BOVIK, A. *The Essential Guide to Image Processing*. Elsevier Science, 2009. ISBN 9780080922515. Disponível em: <https://books.google.com.br/books?id=6TOUgytafmQC>. Citado na página 19.

BRACEWELL, R. N. *The Fourier Transform and Its Applications*. McGraw Hill, 2000. (Electrical engineering series). ISBN 9780073039381. Disponível em: <https://books.google.com.br/books?id=ZNQQAQAAIAAJ>. Citado na página 7.

BRITANAK, V.; YIP, P.; RAO, K. *Discrete Cosine and Sine Transforms: General Properties, Fast Algorithms and Integer Approximations*. Elsevier Science, 2010. ISBN 9780080464640. Disponível em: <https://books.google.com.br/books?id=iRlQHcK-r_kC>. Citado na página 8.

CodeSignal. Imagem, *Arquitetura de uma rede neural perceptron multicamadas (MLP)*. n.d. Acesso em: 19 dez. 2025. Disponível em: <https://codesignal.com/learn/courses/the-mlp-architecture-activations-initialization/lessons/stacking-layers-building-a-multi-layer-perceptron-mlp-1>. Citado na página 16.

DONG, C. et al. Compression artifacts reduction by a deep convolutional network. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. p. 576–584. Citado na página 1.

DONG, C. et al. Learning a deep convolutional network for image super-resolution. In: FLEET, D. et al. (Ed.). *Computer Vision – ECCV 2014*. Cham: Springer International Publishing, 2014. p. 184–199. ISBN 978-3-319-10593-2. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-319-10593-2_13>. Citado na página 1.

DONTRE, A. J. Resolution reassurance for video quality of experience: Attribute substitution in a new context. *Available at SSRN 4446577*, 2023. Disponível em: <http://dx.doi.org/10.2139/ssrn.4446577>. Citado na página 1.

ELFWING, S.; UCHIBE, E.; DOYA, K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *Neural Networks*, v. 107, p. 3–11, 2018. ISSN 0893-6080. Special issue on deep reinforcement learning. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0893608017302976>. Citado na página 17.

FRASSON, M. V. S. *Transformada Discreta de Cosseno: uma aplicação da Álgebra Linear na compressão de imagens do formato JPEG*. [S.l.], 2013. ICMC-USP. Acesso em: Agosto de 2024. Disponível em: <https://sites.icmc.usp.br/frasson/jpeg/jpeg.pdf>. Citado na página 9.

GALL, D. L. Mpeg: a video compression standard for multimedia applications. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 34, n. 4, p. 46–58, abr. 1991. ISSN 0001-0782. Disponível em: <https://doi.org/10.1145/103085.103090>. Citado na página 1.

GLOROT, X.; BENGIO, Y. Understanding the difficulty of training deep feedforward neural networks. In: TEH, Y. W.; TITTERINGTON, M. (Ed.). *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*. Chia Laguna Resort, Sardinia, Italy: PMLR, 2010. (Proceedings of Machine Learning Research, v. 9), p. 249–256. Disponível em: <https://proceedings.mlr.press/v9/glorot10a.html>. Citado na página 29.

GONZALEZ, R.; WOODS, R. *Digital Image Processing, Global Edition*. Pearson Education, 2018. ISBN 9781292223070. Disponível em: <https://books.google.com.br/books?id=P8AoEAAAQBAJ>. Citado 2 vezes nas páginas 5 e 19.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep Learning*. [S.l.]: MIT Press, 2016. <http://www.deeplearningbook.org>. Citado na página 1.

GUNAWAN, R. et al. Optimizing natural image quality evaluators for quality measurement in ct scan denoising. *Computers*, v. 14, n. 1, 2025. ISSN 2073-431X. Acesso em: Outubro de 2025. Disponível em: <https://www.mdpi.com/2073-431X/14/1/18>. Citado na página 20.

HAMILTON, E. *JPEG File Interchange Format*. [S.l.], 1992. Acesso em: Agosto de 2024. Disponível em: <https://www.w3.org/Graphics/JPEG/jfif3.pdf>. Citado na página 6.

HAUTIèRE, N.; TAREL, J.-P.; BRéMOND, R. Perceptual hysteresis thresholding: Towards driver visibility descriptors. In: . [S.l.: s.n.], 2007. p. 89 – 96. ISBN 978-1-4244-1491-8. Citado na página 6.

HAYKIN, S. *Neural Networks and Learning Machines*. Prentice Hall, 2009. (Neural networks and learning machines, v. 10). ISBN 9780131471399. Disponível em: <https://books.google.com.br/books?id=K7P36lKzI_QC>. Citado na página 15.

HE, K. et al. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778. Citado 2 vezes nas páginas 18 e 29.

HENDRYCKS, D.; GIMPEL, K. *Gaussian Error Linear Units (GELUs)*. 2023. Disponível em: <https://arxiv.org/abs/1606.08415>. Citado na página 17.

HUFFMAN, D. A. A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, v. 40, n. 9, p. 1098–1101, 1952. Citado na página 13.

HWANG, J. *Multimedia Networking: From Theory to Practice*. Cambridge University Press, 2009. ISBN 9780521882040. Disponível em: <https://books.google.com.br/books?id=5q2g7LkcNSkC>. Citado 2 vezes nas páginas 1 e 19.

IBM. *What is Gradient Descent?* 2025. IBM Think - The 2025 Guide to Machine Learning. Acesso em: 10 nov. 2025. Disponível em: <https://www.ibm.com/think/topics/gradient-descent>. Citado na página 17.

ITU-T. *Information technology - Digital compression and coding of continuous-tone still images – Requirements and guidelines*. [S.l.], 1992. ISO/IEC 10918-1. Acesso em: Agosto de 2024. Disponível em: <https://www.itu.int/rec/T-REC-T.81>. Citado 3 vezes nas páginas 4, 12 e 60.

KHRISTOV, A. *Zig-zag pattern used in JPEG entropy coding.* 2007. Acesso em: 22 abr. 2025. Disponível em: <https://commons.wikimedia.org/wiki/File:JPEG_ZigZag.svg>. Citado na página 13.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–44, 05 2015. Citado na página 1.

LIVERA, A.; RIBEIRO, N. M. *Transformada Discreta do Coseno (DCT) - Universidade Fernando Pessoa*. 2014. Acesso em: Novembro de 2024. Disponível em: <http://multimedia.ufp.pt/codecs/compressao-com-perdas/metodos-baseados-em-transformadas/transformada-discreta-do-coseno-dct/>. Citado na página 8.

MITTAL, A.; SOUNDARARAJAN, R.; BOVIK, A. C. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, v. 20, n. 3, p. 209–212, 2013. Citado 2 vezes nas páginas 2 e 21.

MONTGOMERY, C. M. *next generation video: Introducing Daala*. 2013. Acesso em: 17 dez. 2025. Disponível em: <https://people.xiph.org/~xiphmont/demo/daala/demo1.shtml>. Citado na página 10.

OLIVEIRA, J. C. de. Compressão, padrões de compressão. In: *Disciplina de Sistemas Multimídia*. Rio de Janeiro, RJ: IME, 2002. Acesso em: 10 nov. 2025. Disponível em: <https://www.lncc.br/~jauvane/SMM/2002/SMM-A04.pdf>. Citado na página 12.

RAO, K. R.; YIP, P. *Discrete Cosine Transform: Algorithms, Advantages, Applications*. San Diego, CA, USA: Academic Press, 1990. ISBN 9780125802031. Citado na página 8.

RUSS, J. C. *The Image Processing Handbook*. [S.l.]: CRC Press, 6ed., 2011. Citado na página 1.

SALOMON, D. *Data Compression: The Complete Reference*. [S.l.: s.n.], 2007. ISBN 978-1-84628-602-5. Citado na página 8.

SHANNON, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, v. 27, n. 3, p. 379–423, 1948. Citado na página 21.

Stanford University. *Lossless Compression of Quantized Values*. n.d. Acesso em: 19 dez. 2025. Disponível em: <https://cs.stanford.edu/people/eroberts/courses/soco/projects/data-compression/lossy/jpeg/lossless.htm>. Citado na página 13.

STOLFI, G. *COMPRESSÃO DE IMAGENS: PADRÃO JPEG*. [S.l.], 2016. Acesso em: 8 nov. 2025. Disponível em: <http://www.lcs.poli.usp.br/~gstolfi/PPT/APTV0616.pdf>. Citado 2 vezes nas páginas 4 e 8.

TECHNOLOGIES, A. *Measuring Video Quality and Performance: Best Practices*. Akamai Technologies, 2020. Acesso em: 13 out. 2025. Disponível em: <http://akamai.com/resources/white-paper/measuring-video-quality-and-performance-best-practices>. Citado na página 1.

TIANGOLO, S. R. *FastAPI*. 2025. A modern, fast (high-performance) web framework for building APIs with Python 3.7+ based on standard Python type hints. Acesso em: 16 nov. 2025. Disponível em: <https://fastapi.tiangolo.com/>. Citado na página 46.

TSAI, D.-Y. et al. Information-entropy measure for evaluation of image quality. *Journal of Digital Imaging*, v. 21, 09 2008. Citado na página 2.

VEMULAPATI, P. *Image Padding Techniques: Reflect Padding (part 2)*. 2023. Acesso em: 15 nov. 2025. Disponível em: <https://medium.com/@Orca_Thunder/image-padding-techniques-reflect-padding-part-2-5a013cd96537>. Citado na página 27.

WALLACE, G. The jpeg still picture compression standard. *IEEE Transactions on Consumer Electronics*, v. 38, n. 1, p. xviii–xxxiv, 1992. Citado 3 vezes nas páginas 1, 3 e 4.

WANG, Z. et al. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, v. 13, n. 4, p. 600–612, 2004. Citado na página 20.

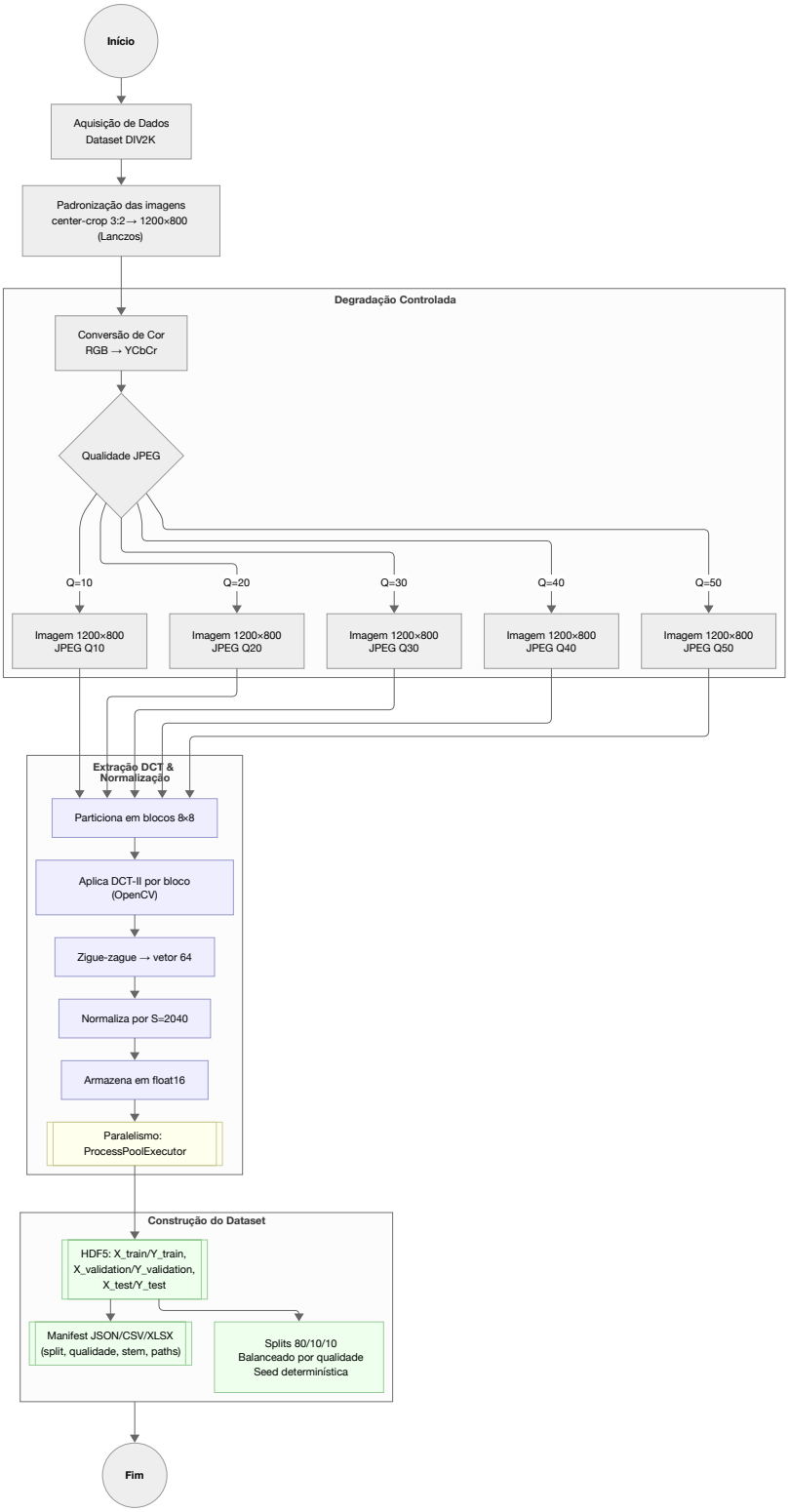
WANG, Z.; SIMONCELLI, E.; BOVIK, A. Multiscale structural similarity for image quality assessment. In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. [S.l.: s.n.], 2003. v. 2, p. 1398–1402 Vol.2. Citado 2 vezes nas páginas 20 e 21.

Wikipedia contributors. *Ringing artifacts — Wikipedia, The Free Encyclopedia*. 2025. [Online; Acesso em: 12 nov. 2025.]. Disponível em: <https://en.wikipedia.org/w/index.php?title=Ringing_artifacts&oldid=1319263253>. Citado na página 14.

XU, J. et al. Understanding and improving layer normalization. In: *Proceedings of the 33rd International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2019. Citado na página 18.

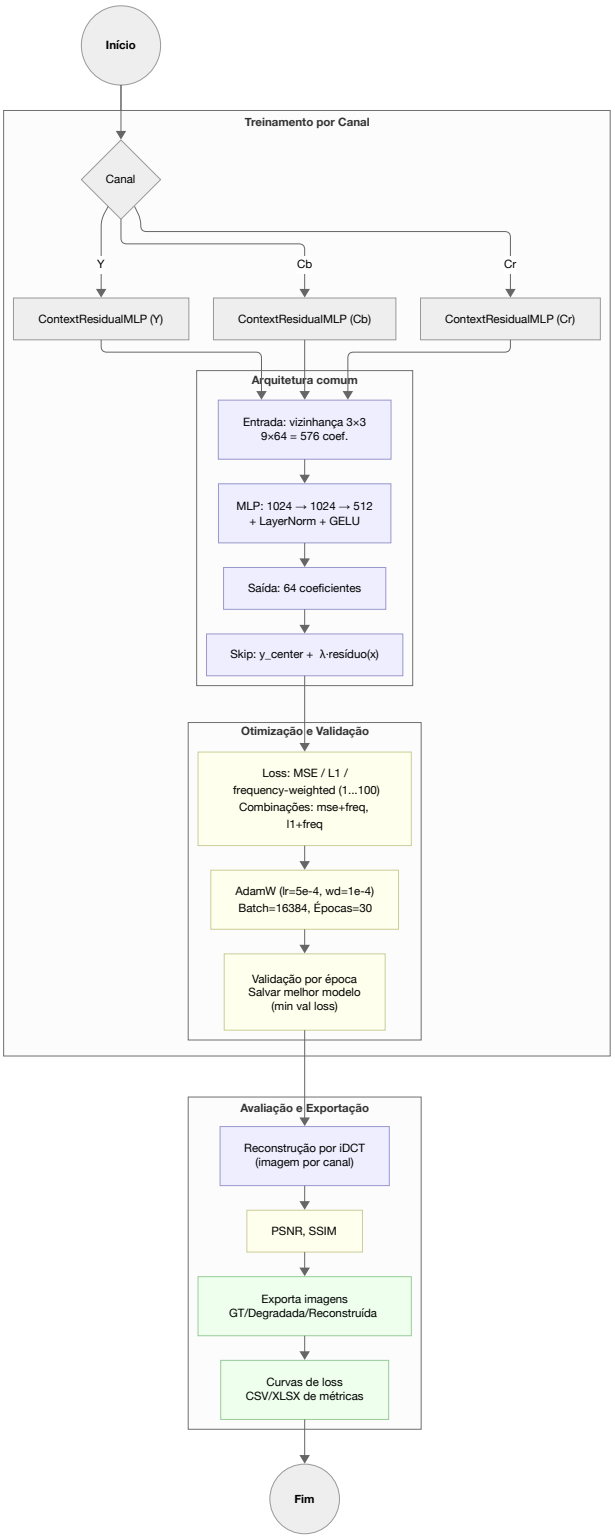
ZAMIR, S. W. et al. Restormer: Efficient transformer for high-resolution image restoration. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2022. p. 5718–5729. Citado na página 1.

APÊNDICE A – DIAGRAMA: PROCESSAMENTO DE DADOS



Fonte: Do próprio autor.

APÊNDICE B – DIAGRAMA: TREINAMENTO DOS MODELOS



Fonte: Do próprio autor.

APÊNDICE C – RESULTADOS COMPLEMENTARES DAS FUNÇÕES DE PERDA

A seguir, são apresentados os resultados complementares da comparação entre as diferentes funções de perda avaliadas, conforme a subseção 3.7.3, incluindo as métricas PSNR, MS-SSIM, NIQE e Entropia. Os valores correspondem à média de 18 imagens para cada nível de qualidade JPEG.

Tabela 7 – Resultados médios de PSNR (dB) e MS-SSIM no canal Y para diferentes funções de perda.

Q	PSNR (dB)					MS-SSIM				
	L1	**L1+Freq**	**MSE**	**MSE+Freq**	**Freq**	**L1**	**L1+Freq**	**MSE**	**MSE+Freq**	**Freq**
10	**29,47**	29,44	28,90	28,78	28,61	**0,9614**	0,9606	0,9561	0,9522	0,9457
20	**32,20**	32,18	31,62	31,53	31,40	**0,9810**	0,9808	0,9791	0,9778	0,9756
30	**32,71**	32,70	32,06	32,02	31,96	**0,9871**	0,9871	0,9856	0,9852	0,9842
40	**33,88**	33,87	33,31	33,27	33,21	**0,9891**	0,9891	0,9879	0,9877	0,9872
50	**33,57**	33,55	33,01	32,99	32,97	**0,9917**	0,9917	0,9908	0,9907	0,9904

Tabela 8 – Resultados médios das métricas sem referência NIQE e Entropia no canal Y para diferentes funções de perda. Valores menores de NIQE indicam maior naturalidade da imagem.

Q	NIQE					Entropia				
	L1	**L1+Freq**	**MSE**	**MSE+Freq**	**Freq**	**L1**	**L1+Freq**	**MSE**	**MSE+Freq**	**Freq**
10	4,73	4,69	4,54	4,57	**4,47**	**7,06**	7,00	6,99	6,88	6,79
20	4,09	4,11	3,94	3,92	**3,78**	**7,27**	7,24	7,26	7,22	7,20
30	3,75	3,76	3,56	3,57	**3,55**	7,34	7,33	**7,36**	7,35	7,35
40	3,57	3,53	3,41	3,43	**3,41**	7,28	7,27	**7,29**	7,28	7,29
50	3,30	3,29	3,13	3,13	**3,11**	7,57	7,57	7,58	7,58	**7,58**

Anexos

ANEXO A – TABELAS DE QUANTIZAÇÃO JPEG

A seguir, apresentam-se as Tabelas de Quantização padrão do JPEG (Annex K), utilizadas para os componentes de luminância e crominância. Essas tabelas foram determinadas com base em experimentos psicovisuais e derivadas de forma empírica.

Segundo o ITU-T (1992, p. 143), se essas tabelas forem divididas por dois — isto é, se seus valores forem reduzidos pela metade, a imagem reconstruída resultante geralmente permanece praticamente indistinguível da imagem original, evidenciando a tolerância do sistema visual humano a quantizações mais agressivas em frequências elevadas.

Tabela 9 – Tabela de quantização JPEG Annex K para luminância

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Tabela 10 – Tabela de quantização JPEG Annex K para crominância

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99