



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE INFORMÁTICA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA COMPUTAÇÃO

JAIR GALVÃO DE ARAÚJO

Sequências de Baixa Discrepância na Aceleração da Avaliação da Qualidade de Imagem
Digital

Recife

2022

JAIR GALVÃO DE ARAÚJO

Sequências de Baixa Discrepância na Aceleração da Avaliação da Qualidade de Imagem
Digital

Tese de Doutorado apresentada ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Doutor em Ciência da Computação.

Área de Concentração: Inteligência Computacional

Orientador: Tsang Ing Ren

Coorientador: Francisco Madeiro Bernardino Junior

Recife

2022

Catálogo na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

A663s Araújo, Jair Galvão de
Sequências de baixa discrepância na aceleração da avaliação da qualidade de imagem digital / Jair Galvão de Araújo. – 2022.
119 f.: il., fig.

Orientador: Tsang Ing Ren.
Tese (Doutorado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2022.
Inclui referências e anexo.

1. Inteligência computacional. 2. Qualidade de imagem - avaliação. I. Ren, Tsang Ing (orientador). II. Título.

006.31

CDD (23. ed.)

UFPE - CCEN 2023-11

Jair Galvão de Araújo

“Sequências de Baixa Discrepância na Aceleração da Avaliação da Qualidade de Imagem Digital”

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Ciência da Computação. Área de Concentração: Inteligência Computacional.

Aprovado em: 12/09/2022.

Orientador: Prof. Dr. Tsang Ing Ren

BANCA EXAMINADORA

Prof. Dr. Carlos Alexandre Barros de Mello
Centro de Informática / UFPE

Prof. Dr. George Darmiton da Cunha Cavalcanti
Centro de Informática / UFPE

Prof. Dr. Cleber Zanchettin
Centro de Informática / UFPE

Profa. Dra. Mylène Christine Queiroz de Farias
Departamento de Engenharia Elétrica / UnB

Prof. Dr. Ricardo da Silva Torres
Department of ICT and Natural Sciences /
Norwegian University of Science and Technology

Dedico este trabalho a minha mãe, Cássia Galvão da Silva, minha namorada, Jessika Gabriela Celestino de Souza, e aos professores Tsang e Madeiro.

AGRADECIMENTOS

Agradeço ao meu orientador, Professor Tsang, por toda a atenção e o apoio para a realização deste trabalho. Levarei todo aprendizado vivido nesta jornada. Muito obrigado.

Agradecimentos ao Professor Madeiro pelo suporte e co-orientação no andamento deste trabalho.

Agradeço a meus familiares, que sempre confiaram em mim e foram de extrema importância para a realização deste trabalho. Obrigado mãe, irmão e minha namorada.

Agradeço aos meus amigos de trabalho e professores do Instituto Federal de Pernambuco e do Instituto Federal do Sertão Pernambucano.

Agradeço aos professores participantes da banca por aceitarem o convite da minha defesa de Tese.

Agradecimentos aos amigos que fiz no grupo de pesquisa VIISAR, pelas excelentes reuniões científicas.

Agradeço aos servidores da secretaria da Pós-graduação do Centro de informática - UFPE.

RESUMO

O crescente uso de dispositivos móveis e aplicações faz surgir a preocupação com a qualidade de imagens e vídeos produzidos, compartilhados e armazenados. Neste cenário, a Avaliação da Qualidade de Imagem, *Image Quality Assessment* (IQA), desempenha um papel importante no processamento de imagens digitais para o fornecimento de serviços satisfatórios aos usuários. A Avaliação da Qualidade de Imagem pode ser dividida em avaliação subjetiva e objetiva. A IQA subjetiva obtém o julgamento através da apresentação da imagem ao Sistema Visual Humano, já a IQA objetiva busca reproduzir a avaliação subjetiva. No entanto, os algoritmos IQA objetivos podem exigir grande número de operações matemáticas para realizar a avaliação. Além disso, os algoritmos IQA objetivos baseados em aprendizagem de máquina têm um elevado custo computacional na fase de treinamento. Nesta tese é proposto um *framework* para avaliação da qualidade de imagens digitais utilizando uma estratégia de amostragem de blocos de imagens e remoção da aleatoriedade na coleta dos pixels sobre a cobertura da imagem. Primeiramente, a estratégia de amostragem com pontos de baixa discrepância foi utilizada a partir das sequências Van der Corput-Halton, Sobol e amostragem Uniforme para coletar blocos de pixels da imagem para formar uma representação com menor número de pixels. Em seguida, essas imagens são avaliadas por métodos de IQA com ou sem a presença da imagem de referência. Os índices de qualidade das imagens amostradas foram comparadas com os índices obtidos das imagens com transformações *Resize* e *Cropped* contendo a mesma quantidade de pixels. Os resultados dos experimentos revelam que a estratégia de amostragem é suficiente para obter as mesmas correlações de qualidade em relação à avaliação subjetiva ao utilizar aproximadamente 8% dos pixels das imagens da base Live e 7% da TID2013 quando comparado com as correlações obtidas das imagens com todos os pixels pelas métricas PSNR e SSIM. Outrossim, o método Learned Perceptual Image Patch Similarity (LPIPS), treinado com imagens amostradas a partir de pontos da sequência VDH, obtém 68,30% de taxa de acerto sobre as opiniões subjetivas de qualidade na base Berkeley-Adobe Perceptual Patch Similarity (BAPPS). A economia de tempo no treinamento do método *Deep Image Quality Measure for FR* (DIQaM-FR) foi 3,61 vezes, diminuição de 27,56 horas para 8,6 horas, para a base Live.

Palavras-chaves: avaliação da qualidade de imagem; sequência de baixa discrepância;

amostragem de imagem digital.

ABSTRACT

The growing use of mobile devices and applications raises concerns about the quality of images and videos produced, shared, and stored. In this scenario, Image Quality Assessment (IQA) plays an essential role in digital image processing to provide satisfactory services to users. The Image Quality Assessment can be divided into subjective and objective assessments. The subjective IQA obtains the judgment by presenting the image to the Human Visual System (HVS) while the objective IQA seeks to reproduce the subjective evaluation. However, objective IQA algorithms may require many mathematical operations to perform the assessment. Furthermore, objective IQA algorithms based on machine learning have a high computational cost in the training phase. This thesis proposes a framework for evaluating the quality of digital images using a sampling strategy of blocks of images and removing randomness in collecting pixels over the image coverage. First, the sampling strategy with low discrepancy points was used from the Van der Corput-Halton, Sobol, and Uniform sampling sequences to collect blocks of pixels from the image to form a representation with fewer pixels. Then, these images are evaluated by WQI methods with or without the presence of the reference image. The quality indexes of the sampled images were compared with the indexes obtained from images with resized and cropped transformations containing the same amount of pixels. Experimental results reveal that the sampling strategy is sufficient to get the same quality correlations concerning the subjective evaluation when using approximately 8% of the pixels of the Live base images and 7% of the TID2013 when compared with the correlations obtained from the images with all pixels by PSNR and SSIM metrics. Furthermore, the Learned Perceptual Image Patch Similarity (LPIPS) method, trained with images sampled from points in the VDH sequence, obtains a 68.30% accuracy rate on subjective quality opinions in the Berkeley-Adobe Perceptual Patch Similarity (BAPPS) database.). The time savings in training the Deep Image Quality Measure for FR (DIQaM-FR) method was 3.61 times, decreasing from 27.56 hours to 8.6 hours for the Live base.

Keywords: image quality assessment; low-discrepancy sequences; digital image sampling.

LISTA DE FIGURAS

Figura 1 – Exemplo de aplicação que utiliza o índice de qualidade como uma etapa de um sistema de melhoramento da qualidade de imagem degradada.	32
Figura 2 – Estrutura de um <i>framework</i> utilizado por métodos de Avaliação da Qualidade de referência completa.	32
Figura 3 – Modelos do gradiente Roberts.	36
Figura 4 – Modelo dos métodos <i>Deep Image Quality Measure for Full Reference Image Quality Assessment</i> (DIQaM-FR) e <i>WaDIQaM-FR</i> (BOSSE et al., 2018).	41
Figura 5 – Diagrama do funcionamento do método <i>Learned Perceptual Image Patch Similarity</i> (LPIPS) composto por duas etapas: cálculo da distância e a predição do julgamento perceptivo. Para calcular a distância d_0 entre dois blocos x, x_0 em uma rede convolucional F , primeiramente, os atributos da rede F são extraídos e são normalizados. As ativações por canal são escaladas pelo vetor W e calculada a distância l_2 . Por fim, a média espacial é calculada e somada por canal.	42
Figura 6 – Visão geral do framework Blinds-II.	43
Figura 7 – Diagrama de blocos do método <i>Perception-based Image Quality Evaluator</i> (PIQUE).	48
Figura 8 – Exemplo de grade de filtros ($w \times w$) de tamanho $w = 3$	51
Figura 9 – 100 pontos aleatórios preenchendo todo o intervalo $[0, 1]$	53
Figura 10 – Gráficos de dispersão de 256 pontos bidimensionais no intervalo $[0, 1]$ obtidos das sequências Sobol, VDH, Uniforme e Aleatória. Os eixos horizontal e vertical são coordenadas dos pontos.	55
Figura 11 – Exemplo de amostragem com 10 pontos gerados a partir das sequências Sobol e Aleatória.	57
Figura 12 – Estratégia de amostragem para avaliação da qualidade de imagens baseada em amostragem por blocos de pixel. A estratégia inicia com conjuntos de pontos que correspondem a um bloco na imagem. Em seguida, os blocos são extraídos, concatenados e apresentados à avaliação da qualidade.	59

Figura 13 – Exmplo de extração de 64 blocos de tamanho de 16×16 pixels sobre a imagem I23 da base TID2013. Foram utilizados 64 pontos da sequência Van der Corput-Halton (VDH) mapeados sobre o espaço de 32×24 blocos de pixels da imagem.	61
Figura 14 – Exemplos de concatenação lado a lado e vertical.	62
Figura 15 – Exemplo de transformações da imagem I22 de tamanho 512×384 pixels da base TID2013. O redimensionamento e o recorte ficaram com 175×175 pixels após as transformações.	64
Figura 16 – Imagens de referência utilizadas nas simulações da base de imagens Live.	66
Figura 17 – Imagens de referência utilizadas nas simulações da base de imagens TID2013.	67
Figura 18 – Correlação SROCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatória variando o tamanho do bloco ($2^M \times 2^M$) na base de imagem Live. As imagens amostradas foram avaliadas pelo Structural Similarity.	72
Figura 19 – Correlação LCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatório variando o tamanho do bloco ($2^M \times 2^M$) na base de imagem Live. As imagens amostradas foram avaliadas pelo Structural Similarity.	72
Figura 20 – Análise da correlação SROCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatório variando o tamanho do bloco ($2^M \times 2^M$) nas imagens da base Live. As imagens amostradas foram avaliadas pelo método DIQaM-FR treinado com 1.000 épocas.	81
Figura 21 – Análise da corralação LCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatório variando o tamanho do bloco ($2^M \times 2^M$) nas imagens do banco de dados Live. As imagens amostradas foram avaliadas pelo método DIQaM-FR treinado com 1.000 épocas.	82
Figura 22 – A taxa de sucesso dos métodos de amostragem no julgamento da escolha forçada de duas alternativas (2AFC) com LPIPS em imagens de real algoritmos (CNN e distorções tradicionais).	89
Figura 23 – A taxa de sucesso de métodos de amostragem em julgamento de escolha forçada de duas alternativas (2AFC) com LPIPS em imagens de distorções (super-resolução, interpolação de quadros, desfoque e colorização).	89

Figura 24 – Análise do histograma das imagens Student Sculpture e Dancers do banco de imagens Live. As imagens foram amostradas pelas sequências VDH, Sobol e Uniforme. A taxa de amostragem foi de 30 blocos de 32×32 pixels.	90
Figura 25 – Análise do histograma das imagens I13 e I15 do banco de imagens TID2013. As imagens foram amostradas pelas sequências VDH, Sobol e Uniforme. A taxa de amostragem foi de 12 blocos de 32×32 pixels.	90
Figura 26 – Comparativo do desempenho das correlações SROCC e LCC das métricas IQA obtidos das imagens com todos os pixels e amostradas. A métrica LPIPS avaliou as imagens da base BAPPS e as métricas PSNR, SSIM, Fast-SSIM, BRISQUE, NIQE e PIQUE avaliaram a base Live. O quantitativo no topo das correlações apresenta a aceleração obtida com a estratégia de amostragem. A Figura 26a teve seu tempo medido em segundos e a Figura 26b teve o tempo de treinamento das métricas medido em horas.	101
Figura 27 – Comparativo do desempenho das correlações SROCC e LCC das métricas IQA obtidos das imagens com todos os pixels e amostradas da base TID2013. O quantitativo no topo das correlações apresenta a aceleração obtida com a estratégia de amostragem. A Figura 26a teve seu tempo medido em segundos e a Figura 26b teve o tempo de treinamento das métricas medido em horas.	101
Figura 28 – Exemplo de um quadro do conteúdo de vídeo 5 com sequência 1 do banco de imagens LIVE-Netflix (BAMPIS et al., 2016).	103

LISTA DE TABELAS

Tabela 1 – Escala categórica para ser utilizada na avaliação subjetiva de imagens digitais.	31
Tabela 2 – Exemplo dos primeiros números da sequência VDH obtidos pela reversão (R) dos números da base binária.	58
Tabela 3 – Mapeamento de um ponto bidimensional para o espacial de uma imagem com dimensões de 512×384 pixels. As 1° e 2° dimensões do ponto são mapeadas para as dimensões da imagem.	60
Tabela 4 – Quantitativo de correspondência do valor de SSIM usando amostragem pixel a pixel em relação ao valor de SSIM avaliando as imagens da base Live (total de 982 imagens) com todos os pixels.	69
Tabela 5 – Avaliação das correlações SROCC, LCC e RMSE da amostragem pixel a pixel utilizando o método Structural Similarity aplicado à base Live. O μ denota a média dos valores de SSIM das imagens avaliadas, σ denota o desvio padrão e p corresponde ao p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.	70
Tabela 6 – Valores médios de SSIM e correlações SROCC e LCC a partir das imagens amostradas usando sequências de baixa discrepância sobre as bases Live (30 blocos) e TID2013 (12 blocos). As imagens foram amostradas com blocos de tamanhos de 32×32 pixels. O μ denota a média dos valores do SSIM, σ denota o desvio padrão e p corresponde p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.	74
Tabela 7 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e avaliadas pelo método SSIM. O teste de Wilcoxon foi aplicado. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferença significativa entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com a estratégia de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	75

Tabela 8 – O número de imagens com o valor SSIM dentro do intervalo de até 8% de diferença do valor SSIM original usando uma porcentagem de pixels das imagens na base Live, total de 982 imagens.	75
Tabela 9 – O número de imagens com o valor SSIM dentro do intervalo de até 8% de diferença do valor SSIM original usando uma porcentagem de pixels das imagens na base TID2013, total de 3000 imagens.	75
Tabela 10 – Valores médios de <i>Peak Signal-to-Noise Ratio</i> (PSNR) e correlações SROCC e LCC das imagens amostradas usando sequências de baixa discrepância sobre a base Live e TID2013. As imagens foram amostradas com 30 blocos para Live e 12 blocos para a TID2013, o tamanho do bloco foi de 32×32 pixels. O μ denota a média dos valores do PSNR, σ denota o desvio padrão e p corresponde p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.	76
Tabela 11 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e avaliadas pelo método PSNR. O teste aplicado foi o teste de Wilcoxon. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	77
Tabela 12 – Valores médios de <i>Fast Structural Similarity</i> (Fast-SSIM) e correlações SROCC e LCC a partir das imagens amostradas usando sequências de baixa discrepância sobre a base Live. As imagens foram amostradas com 30 blocos de 32×32 pixels. O μ denota a média dos valores do Fast-SSIM, σ denota o desvio padrão e p corresponde p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.	78

Tabela 13 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações <i>Resize</i> e <i>Cropped</i> e avaliadas pelo método Fast-SSIM. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	79
Tabela 14 – Variação do tamanho do bloco, número de blocos e camadas convolucionais para cada tamanho de bloco na etapa de treinamento.	80
Tabela 15 – Desempenho do modelo DIQaM-FR com duas estratégias de <i>pooling</i> . O tamanho do bloco foi de 4×4 pixels no banco de dados Live. O modelo foi treinado com imagens de amostragem e testado em imagens de amostragem e imagens completas.	83
Tabela 16 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações <i>Resize</i> e <i>Cropped</i> e avaliadas pelo método DIQaM-FR. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	84

Tabela 17 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações <i>Resize</i> e <i>Cropped</i> e avaliadas pelo método WaDIQaM-FR. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	84
Tabela 18 – Desempenho do modelo DIQaM-FR com duas estratégias de <i>pooling</i> . O tamanho do bloco foi de 4×4 pixels no banco de dados TID2013. O modelo foi treinado com imagens de amostragem e testado em imagens de amostragem e imagens completas.	85
Tabela 19 – Tempo em horas de treinamento gasto para os modelos DIQaM-FR e WaDIQaM-FR.	85
Tabela 20 – A taxa de avaliação perceptual correta do LPIPS treinado sobre as imagens distorcidas a partir da variação dos parâmetros CNN e testado nos bancos de dados de imagens de validação CNN, colorização, Tradicional, Desfocamento, Super-resolução e Interpolação de quadro. Usamos o μ para denotar a taxa média de acertos e σ para o desvio padrão.	87
Tabela 21 – Correlações SROCC e LCC do LPIPS sobre as imagens distorcidas a partir da variação dos parâmetros CNN e testado nos bancos de dados de imagens de validação CNN, colorização, tradicional, desfocamento, Super-resolução e interpolação de quadro	88
Tabela 22 – Análise de correlação dos histogramas obtidos a partir das imagens amostradas Student Sculpture e Dancers da base Live, I13 e I15 da base TID2012. As sequências VDH, Sobol e Uniforme foram utilizadas para amostrar as imagens. A taxa de amostragem para as imagens da base Live foi de 30 blocos e para as imagens da base TID2013 foi de 12 blocos com cada bloco de tamanho de 32×32 pixels.	91

Tabela 23 – Spearman Rank-Order Correlation (SROCC) e Linear Correlation Coefficient (LCC) a partir de estratégias de amostragem aplicadas o método BRISQUE nas bases de dados Live e TID2013. Usamos o p para denotar o valor-p do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.	93
Tabela 24 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações <i>Resize</i> e <i>Cropped</i> e avaliadas pelo método BRISQUE. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	93
Tabela 25 – As correlações Spearman Rank-Order Correlation (SROCC) e Linear Correlation Coefficient (LCC) são apresentadas a partir da avaliação das imagens amostradas pelo método NIQE nas bases Live e TID2013. O p denota o valor-p do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.	95
Tabela 26 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações <i>Resize</i> e <i>Cropped</i> e avaliadas pelo método NIQE. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	95

Tabela 27 – As correlações Spearman Rank-Order Correlation (SROCC) e Linear Correlation Coefficient (LCC) são apresentadas a partir da avaliação das imagens amostradas pelo método PIQUE nas bases Live e TID2013 com 100 blocos de 16×16 e 32×32 pixels, respectivamente. O p denota o valor-p do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.	96
Tabela 28 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações <i>Resize</i> e <i>Cropped</i> e avaliadas pelo método PIQUE. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.	97
Tabela 29 – Valores de SSIM obtidos pelo método de Bruni <i>et al.</i> e das imagens com a estratégia de amostragem utilizando as sequências Sobol, VDH e Uniforme. As simulações foram aplicados na mesma imagem I16 com seus respectivas distorções do banco de imagens TID2013. $\Delta_{x,y} = x - y $.	99
Tabela 30 – Número de operações por pixel (Nopp) necessárias por métrica para avaliar uma imagem de 512×384 pixels.	99
Tabela 31 – Tempos gastos por método de IQA na avaliação das imagens com todos os pixels e com a estratégia de amostragem.	102
Tabela 32 – A comparação de velocidade de avaliação do SSIM e do Fast-SSIM usou uma taxa de amostragem dos blocos de 35 de 50×50 pixels sobre o vídeo 5 e na sequência 1 do banco de dados Live Netflix. O vídeo progressivo tem 250 quadros com 1920×1080 pixels de resolução. . . .	104

LISTA DE ABREVIATURAS E SIGLAS

2AFC	<i>Judgments From Two Alternatives Forced Choices</i>
BAPPS	<i>Berkeley-Adobe Perceptual Patch Similarity</i>
BIBE	<i>Blind Image Blur Quality Evaluator</i>
BLIINDS-II	<i>Blind Image Integrity Notator Using DCT Statistics</i>
BRISQUE	<i>Blind/Referenceless Image Spatial Quality Evaluator</i>
BSIF	<i>Binarized Statistical Image Features</i>
CLBP	<i>Completed Local Binary Patterns</i>
CNN	<i>Convolutional Neural Network</i>
CORNIA	<i>Codebook Representation for No-Reference Image Assessment</i>
CSF	<i>Contrast Sensitivity Function</i>
DCT	Discrete Cosine Transform
DIIVINE	<i>Distortion Identification-based Image Verity and Integrity Evaluation</i>
DIQA	<i>Deep Image Quality Assessor</i>
DIQaM-FR	<i>Deep Image Quality Measure for Full Reference Image Quality Assessment</i>
DMOS	<i>Difference Mean Opinion Score</i>
ELM	<i>Extreme Learning Machine</i>
ESSIM	<i>Edge Strength Similarity</i>
Fast-SSIM	<i>Fast Structural Similarity</i>
FR	<i>Full Reference</i>
FSIM	<i>Feature Similarity</i>
GBM	<i>Gradient Boosting Machine</i>
GMS	<i>Gradient Magnitude Similarity</i>
GMSD	<i>Gradient Magnitude Similarity Deviation</i>
GP-NR-IQA	<i>General-Purpose No-Reference Image Quality Assessment</i>

HVS	<i>Human Visual System</i>
IL-NIQE	<i>Integrated Local Naturalness Image Quality Evaluator</i>
IQA	<i>Image Quality Assessment</i>
ITU	<i>International Telecommunication Union</i>
IW-SSIM	<i>Combining Information Content Weighting SSIM</i>
KRCC	<i>Kendall Rank Correlation Coefficient</i>
LCC	<i>Linear Correlation Coefficient</i>
LCP	<i>Local Configuration Patterns</i>
LPIPS	<i>Learned Perceptual Image Patch Similarity</i>
LPQ	<i>Local Phase Quantization</i>
LRSN	<i>Siamese Convolutional Neural Networks</i>
MAD	<i>Most Apparent Distortion</i>
MDSI	<i>Mean Deviation Similarity Index</i>
ML	<i>Machine Learning</i>
MLIQM	<i>Machine Learning-based Image Quality Measure</i>
MMF	<i>Multi-method Fusion</i>
MOS	<i>Mean Opinion Score</i>
MS-SSIM	<i>Multi-scale Structural Similarity</i>
MSE	<i>Mean Squared Error</i>
NIQE	<i>Naturalness Image Quality Evaluator</i>
NMF	<i>Non-negative Matrix Factorization</i>
NR	<i>No-Reference</i>
NSS	<i>Natural Scene Statistics</i>
OCPP	<i>Orthogonal Color Planes Pattern</i>
OG-IQA	<i>Oriented Gradients Image Quality Assessment</i>
OSS-SSIM	<i>Optimal Scale Selection SSIM</i>
PIQUE	<i>Perception-based Image Quality Evaluator</i>

PSNR	<i>Peak Signal-to-Noise Ratio</i>
RIQA	<i>Referenceless Image Quality Assessment</i>
RR	<i>Reduced Reference</i>
SCNN	<i>Siamese Convolutional Neural Network</i>
SLFNs	<i>Single-hidden Layer Feedforward Neural Networks</i>
SROCC	<i>Spearman Rank Order Correlation Coefficient</i>
SSEQ	<i>Spatial-spectral Entropy-based Quality</i>
SSIM	<i>Structural Similarity</i>
SVD	<i>Singular Value Decomposition</i>
SVM	<i>Support Vector Machine</i>
VDH	Van der Corput-Halton
VSI	<i>Visual Saliency-based Index</i>
WaDIQaM	<i>Image Quality Pooling by Weighted Average</i>

SUMÁRIO

1	INTRODUÇÃO	23
1.1	MOTIVAÇÃO	25
1.2	QUESTÕES DE PESQUISA	27
1.3	HIPÓTESES DA PESQUISA	28
1.4	OBJETIVOS	28
1.5	ESTRUTURA DA TESE	29
2	FUNDAMENTAÇÃO TEÓRICA	30
2.1	AVALIAÇÃO DA QUALIDADE DE IMAGENS DIGITAIS	30
2.1.1	Avaliação subjetiva	30
2.1.2	Avaliação objetiva	31
2.1.3	Métodos de referência completa	32
2.1.4	Métodos não referenciados	42
2.1.5	Avaliação da qualidade de imagem por amostragem	50
3	SEQUÊNCIAS DE BAIXA DISCREPÂNCIA NA AMOSTRA- GEM DE IMAGENS DIGITAIS	53
3.1	SEQUÊNCIAS ALEATÓRIAS	53
3.2	SEQUÊNCIAS DE BAIXA DISCREPÂNCIA	54
3.2.1	Sequência de Sobol	55
3.2.2	Sequência de Van der Corput-Halton	57
3.3	MÉTODO DE AMOSTRAGEM DE IMAGENS COM SEQUÊNCIAS DE BAIXA DISCREPÂNCIA	59
3.3.1	Conjuntos de pontos	59
3.3.2	Extração dos blocos de pixel	60
3.3.3	Concatenação dos blocos de pixel	62
3.3.4	Avaliação da qualidade das amostras	62
4	EXPERIMENTOS E RESULTADOS	64
4.1	PROTOCOLO DOS EXPERIMENTOS	64
4.2	BANCOS DE IMAGENS	66
4.3	AMOSTRAGEM PIXEL A PIXEL	68
4.4	AMOSTRAGEM EM MÉTODOS DE REFERÊNCIA COMPLETA	71

4.4.1	Estratégia de amostragem no método <i>Structural Similarity</i> . . .	71
4.4.2	Estratégia de amostragem no método <i>Peak Signal-to-Noise Ratio</i>	74
4.4.3	Estratégia de amostragem no método <i>Fast-SSIM</i>	77
4.4.4	Estratégia de amostragem no método <i>Deep Image Quality Measure for Full Reference</i>	79
4.4.5	Estratégia de amostragem no método <i>Learned Perceptual Image Patch Similarity</i>	84
4.5	ANÁLISE DE HISTOGRAMA	89
4.6	AMOSTRAGEM EM MÉTODOS NÃO REFERENCIADOS	91
4.6.1	Estratégia de amostragem no método <i>Blind/Referenceless Image Spatial Quality Evaluator</i>	92
4.6.2	Estratégia de amostragem no método <i>Naturalness Image Quality Evaluator</i>	94
4.6.3	Estratégia de amostragem no método <i>Perception-based Image Quality Evaluator</i>	94
4.7	ANÁLISE DE NÚMERO DE OPERAÇÕES POR PIXEL	97
4.8	ANÁLISE DA ACELERAÇÃO NO TEMPO NA AVALIAÇÃO DE IMAGENS AMOSTRADAS	100
4.9	AVALIAÇÃO DA AMOSTRAGEM EM VÍDEO	103
5	CONCLUSÃO	105
5.1	DISCUSSÕES E LIMITAÇÕES	106
5.2	TRABALHOS FUTUROS	107
5.3	ARTIGO PUBLICADO	108
	REFERÊNCIAS	109
	ANEXO A – NÚMERO DE OPERAÇÕES POR PIXEL . . .	119

1 INTRODUÇÃO

O crescente uso de dispositivos móveis e aplicações faz surgir a preocupação com a qualidade de imagens e vídeos produzidos, compartilhados e armazenados, além da expansão dos serviços de *streams*, também emerge a atenção com a qualidade dos serviços entregues aos usuários finais. Um ponto a destacar é que, durante a transmissão, as imagens podem sofrer degradações devido à complexidade existente no fornecimento dos serviços, levando a uma baixa qualidade das imagens entregues aos usuários (NIU et al., 2019b).

A imagem digital pode ser afetada por vários tipos de distorções, por exemplo, mudança de cor, compressão e embaçamento, os quais podem resultar baixa qualidade da imagem fornecida (MANSOURI; MAHMOUDI-AZNAVEH, 2019). Assim, a qualidade da imagem é definida pela avaliação do conjunto de degradações visuais nela encontradas (KAMBLE; BHURCHANDI, 2015). A quantificação precisa e confiável da qualidade é crucial para uma gama de aplicações em processamento de imagem, computação gráfica e comunicação multimídia, incluindo compressão e restauração (BECKER; WIEGAND; BOSSE, 2021).

A qualidade da imagem digital entregue ao destinatário é um ponto relevante em sistemas de processamento de imagem para fornecimento de serviços de qualidade. Desta forma, a Avaliação da Qualidade de Imagem, *Image Quality Assessment* (IQA), desempenha um papel importante na avaliação da qualidade no processamento de imagens digitais. A IQA pode ser dividida em avaliação subjetiva ou objetiva. A avaliação da qualidade mais precisa dos sinais distorcidos ou modificados são os humanos (avaliação subjetiva). A avaliação subjetiva obtém o julgamento através da apresentação das imagens ao sistema visual humano, *Human Visual System* (HVS). O HVS processa o estímulo percebido do sinal para realizar a avaliação (MANSOURI; MAHMOUDI-AZNAVEH, 2019).

No processo de avaliação subjetiva, as imagens são apresentadas para um conjunto de pessoas e suas opiniões (julgamentos) são coletadas. O julgamento é realizado por teste de preferência ou por pontuação dentro de uma escala de qualidade previamente definida (ITU, 2012). No entanto, a avaliação subjetiva é mais difícil de ser realizada quando comparada à avaliação objetiva, tornando-se um processo custoso que demanda maior tempo no processo de avaliação. Assim, a IQA subjetiva é difícil de aplicar em sistemas que necessitam de uma rápida resposta da qualidade devido ao número de julgadores e

ao processo de avaliação, especialmente em sistemas de tempo real (NIU et al., 2019a).

A avaliação objetiva de imagem busca desenvolver algoritmos computacionais para reproduzir a avaliação subjetiva (KONG; CUI; HOU, 2020). A busca por métodos automatizados para investigar a qualidade de imagens digitais que reflitam a avaliação subjetiva fez surgir a área da Avaliação da Qualidade de Imagem. Assim, a IQA é uma área de intensa pesquisa e investigação (NIU et al., 2019a).

Métricas de IQA objetivas visam “imitar” a percepção da visão humana a partir de algoritmos computacionais com abordagens baseadas em treinamento ou extração de parâmetros estatísticos das imagens avaliadas (KAMBLE; BHURCHANDI, 2015). Portanto, o resultado da avaliação das métricas objetivas deve apresentar alta correlação com julgamentos humanos (SHI et al., 2018; BAI; REIBMAN, 2018; ZHU; ZHOU; XUE, 2018). A IQA objetiva busca fornecer uma avaliação de qualidade a partir de atributos ou estatísticas obtidos durante a avaliação do sinal (LING et al., 2022). O desempenho da avaliação da qualidade objetiva da imagem é medido pela precisão na representação da avaliação subjetiva (KONG; CUI; HOU, 2020).

Desenvolver algoritmos de avaliação da qualidade objetivos confiáveis que avaliam a qualidade da imagem automaticamente e que representem a percepção visual humana atraiu uma atenção significativa (ZHOU et al., 2020). A predição da avaliação da qualidade de imagens digitais é essencial para o funcionamento de sistemas, seja no controle, no acompanhamento e na melhoria da qualidade de imagens e vídeos, durante a transmissão ou no armazenamento (KAMBLE; BHURCHANDI, 2015).

A avaliação objetiva da qualidade de imagem é dividida conforme a disponibilidade de informações da imagem de referência em três categorias: Referência Completa, *Full Reference* (FR) (SHI et al., 2018; KONG; CUI; HOU, 2020; CHEN; BOVIK, 2011), Referência reduzida, *Reduced Reference* (RR) (XIE et al., 2019; ZHANG; PHAN; CHANDLER, 2017) e Sem Referência, *No-Reference* (NR) (PO et al., 2019; LI et al., 2020; ZHOU et al., 2020; WANG; LIL; CHEN, 2021; XIANG et al., 2021). Na abordagem FR para avaliar a imagem testada, é necessário ter a sua versão da imagem em seu estado original (imagem de referência), ou seja, imagem sem qualquer modificação, para avaliação de qualidade (REISENHOFER et al., 2018). A avaliação RR tem apenas uma quantidade limitada de atributos extraídos da imagem de referência ou o que esperar de uma imagem sem distorções no cenário de *Natural Scene Statistics* (NSS) para realizar a avaliação da qualidade (WANG; BOVIK, 2011; FREITAS; AKAMINE; FARIAS, 2018a). Métodos de referência completa e casos de métodos

de referência reduzida precisam de informações da imagem de referência para avaliar a qualidade das imagens e definir um índice de qualidade. Quando a imagem de referência não está disponível, a avaliação NR é feita apenas a partir dos atributos extraídos da imagem avaliada (ZHOU et al., 2020).

Os algoritmos IQA objetivos podem exigir grande número de operações matemáticas para realizar a avaliação e, conseqüentemente, maior tempo de avaliação. Além disso, os algoritmos IQA objetivos que utilizam técnicas de aprendizagem de máquina, *Machine Learning* (ML), exigem maior custo computacional na fase de treinamento. Baseado no alto custo computacional e no tempo gasto no processo da avaliação da qualidade, a estratégia de amostragem de imagem digital a partir de pontos predefinidos, como seqüências de baixa discrepância, apresenta-se como uma alternativa para reduzir a quantidade de pixels e evitar a aleatoriedade na coleta de pixels necessários para realizar a avaliação da qualidade.

1.1 MOTIVAÇÃO

A Avaliação da Qualidade de Imagem é uma área de investigação que mede a qualidade de imagens durante o processo de aquisição, armazenamento e na entrega de serviços aos destinatários (LIU et al., 2020). Esse crescente uso de imagens digitais exige que a avaliação da qualidade seja realizada de forma rápida pelos sistemas computacionais. Exemplos destes serviços entregues por aplicações são: *stream* de vídeo, web conferências, serviços de compartilhamento de mídias em redes sociais e web chamadas.

A transmissão de vídeo e imagens em tempo real precisam de respostas imediatas para avaliar a qualidade dos serviços durante a transmissão. Dois dos principais desafios para a avaliação da qualidade de imagem são: (1) maior resolução de imagens e vídeos, como imagens em *4k ultra-high definition* (UHD) (3840×2160 pixels), que implica elevado número de pixels e, conseqüentemente, demanda mais tempo na avaliação, e (2) a complexidade computacional, ou seja, o número de operações que tende a aumentar à medida que a resolução aumenta.

Diante da exigência de rápida resposta no processo da IQA, a avaliação subjetiva não pode ser aplicada em sistemas *online*, por exemplo, serviços de *stream* de vídeo (BAMPIS et al., 2017). Esta exigência motiva o desenvolvimento de algoritmos objetivos de IQA que forneçam uma avaliação rápida da qualidade do serviços entregues. Uma alternativa para

acelerar a avaliação da qualidade de imagens é reduzir a quantidade de pixels envolvidos no processo de avaliação.

Na literatura, estratégias de amostragem são aplicadas para resolução de problemas envolvendo processamento digital de imagens, como classificação de imagem (RONG; WANG; WANG, 2021), super-resolução de imagens (ZHU; GEIß; SO, 2021; SHEN; ZENG; WEI, 2011; AHUJA; BOSE, 2006). No problema de avaliação da qualidade de imagem, Lima *et al.* (LIMA; MADEIRO, 2009) propuseram uma estratégia de amostragem com pontos de baixa discrepância para amostragem de imagem pixel a pixel para avaliar a qualidade de imagens de referência completa apenas pelo o método erro médio quadrático, *Mean Squared Error* (MSE), onde afirmam que amostrando 7% da imagem conseguiram um erro de 5% do valor do MSE (LIMA; MADEIRO, 2009). Bampis *et al.* propuseram uma estratégia de amostragem aleatória baseada em blocos para estimar o valor dos índices dos métodos PSNR, *Structural Similarity* (SSIM), *Combining Information Content Weighting SSIM* (IW-SSIM) e *Feature Similarity* (FSIM) (BAMPIS; GOODALL; BOVIK, 2016). Bruni *et al.* utilizaram uma amostragem baseada em entropia para acelerar a avaliação de imagens pelo método SSIM, com uma estratégia que permite estimar o valor do SSIM com um erro relativo de 5% (BRUNI; VITULANO, 2017). A avaliação não referenciada de Imagens Omnidirecionais Estereoscópicas (SOIs) foi proposta utilizando amostragem de superfície esférica de blocos (QI *et al.*, 2021). Lv *et al.* propuseram uma estratégia de atributos de *hashing* para avaliação de qualidade de referência reduzida, esta estratégia utiliza amostragem aleatória de blocos para estimar o valor do método PSNR (LV; WANG, 2009).

A amostragem de imagens digitais no problema de avaliação da qualidade de imagens é utilizada para realizar a coleta pixel a pixel ou blocos de pixels da imagem. A coleta pixel a pixel em métodos de IQA baseado na comparação da amplitude do sinal se mostra eficiente (LIMA; MADEIRO, 2009), mas, a amostragem pixel a pixel não foi investigada em métodos onde a correlação da vizinhança dos pixels da imagem é importante para extração de atributos, como por exemplo no método SSIM.

A estratégia aleatória de blocos (BAMPIS *et al.*, 2017) exige um custo para gerar pontos aleatórios, além de gerar novos pontos sempre que uma nova imagem é avaliada, consequentemente, aumentando o custo computacional e tempo de avaliação. A estratégia de amostragem de blocos por entropia realiza uma busca na imagem por blocos para formar um conjunto típico de distorções visuais presente na imagem avaliada (BRUNI; VITULANO,

2017).

Este trabalho propõe um *framework* para avaliação da qualidade de imagens digitais utilizando amostragem de blocos de pixels como alternativa para redução do número de operações por pixel e, conseqüentemente, a redução do tempo de avaliação da qualidade de imagens. A estratégia de amostragem possibilita a utilização de seqüências de pontos predeterminados, como as seqüências de baixa discrepância, para coletar blocos de pixels para obter uma representação da imagem com menor quantidade de pixels no processo de avaliação da qualidade de imagens. A estratégia utilizando seqüências de baixa discrepância, como Van der Corput-Halton (CORPUT, 1935) e Sobol (SOBOL, 1967), e a seqüência Uniforme, têm como objetivo coletar pixels da imagem e formar uma representação da imagem no processo de avaliação da qualidade de imagem. Em seguida, a imagem é avaliada por métodos de IQA com ou sem a presença da imagem de referência. Os índices e correlações de qualidade das imagens amostradas a partir das seqüências de baixa discrepância são comparados com amostragem uniforme e aleatória e com as imagens com transformações que reduzem a quantidade de pixels: *Resize* e *Cropped*, ambas contendo a mesma quantidade de pixel.

1.2 QUESTÕES DE PESQUISA

Esta pesquisa é guiada pelas seguintes questões de pesquisa:

- **A utilização de seqüências de pontos predeterminados para coletar blocos de pixels permite uma boa representação da imagem no processo de avaliação da qualidade?**

Uma amostragem da imagem por blocos de pixels pode não apresentar boa representação das distorções visuais presente na imagem. As seqüências de baixa discrepância têm como característica uma boa cobertura de uma região de interesse, deste modo, a amostragem de blocos a partir de pontos de baixa discrepância pode obter uma representação espacial das imagens avaliadas.

- **A concatenação dos blocos de pixels que foram amostrados para formar uma representação menor da imagem influencia na extração de atributos da imagem utilizada na avaliação da qualidade?**

O agrupamento dos blocos amostrados pode influenciar na extração de atributos da vizinhança dos pixels. Por isso, é importante avaliar a concatenação de blocos lado a lado, vertical e horizontal, além de avaliar a quantidade e o tamanho dos blocos na avaliação da qualidade da imagem.

1.3 HIPÓTESES DA PESQUISA

As seguintes hipóteses integram esta pesquisa:

- A amostragem de imagem por blocos de pixels a partir de sequências de pontos predeterminados, como as sequências de baixa discrepância, pode ser utilizada como alternativa para estimar o índice de qualidade dos métodos de IQA com e sem a imagem de referência quando comparado com a avaliação da imagem com todos os pixels.
- A concatenação dos blocos de pixels amostrados para formar a representação com menor quantidade de pixels da imagem pode influenciar na extração de atributos da imagem utilizada na avaliação da qualidade.

1.4 OBJETIVOS

Este trabalho objetiva propor um *framework* que utiliza a estratégia de amostragem de bloco de pixels a partir de sequências de pontos predeterminados, como as sequências de baixa discrepância, em imagens digitais para redução do número de operações realizadas na avaliação da qualidade de imagem e, conseqüentemente, do tempo gasto na avaliação.

Os objetivos específicos deste trabalho são:

- Propor um *framework* para utilização de sequências de pontos para amostragem de imagem digital para métodos de avaliação com e sem acesso as informações da imagem de referência.
- Investigar a estratégia de amostragem pixel a pixel e por bloco de pixels, sob o ponto de vista de desempenho.

- Avaliar a utilização das sequências de baixa discrepância como alternativa para amostragem de imagens digitais.
- Investigar a concatenação dos blocos de pixels na representação das imagens avaliadas.
- Definir uma estratégia de treinamento de métodos de IQA baseado em aprendizagem com imagens amostradas.
- Avaliar a redução da quantidade pixels utilizados para estimar a qualidade da imagem avaliada.
- Comparar o tempo de avaliação da qualidade da imagem completa (com todos os pixels) e da sua versão amostrada.

1.5 ESTRUTURA DA TESE

Além deste capítulo de Introdução, esta Tese está dividida em quatro outros capítulos adicionais que estão organizados da seguinte forma:

- Capítulo 2: Neste capítulo, os fundamentos da avaliação da qualidade de imagem digital são apresentados, além da revisão da literatura dos métodos objetivos de referência completa e não referenciados.
- Capítulo 3: Este capítulo descreve os conceitos de sequência de baixa discrepância e o *framework* para IQA baseado na estratégia de amostragem de blocos a partir de pontos espacialmente previamente localizados.
- Capítulo 4: Este capítulo apresenta os protocolos dos experimentos realizados e os resultados obtidos.
- Capítulo 5: Neste capítulo, são apresentados a conclusão do trabalho, discussões e trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo apresenta os fundamentos sobre Avaliação da Qualidade de Imagens digitais e o seu estado da arte. São apresentados também os métodos de Avaliação da Qualidade de Imagens utilizados nas simulações.

2.1 AVALIAÇÃO DA QUALIDADE DE IMAGENS DIGITAIS

A Avaliação da Qualidade de Imagem, IQA, desempenha um papel importante em aplicações que utilizam processamento de imagens digitais. Na aquisição, no armazenamento e na transmissão de imagens e vídeos digitais, degradações podem ser inseridas e interferindo diretamente na qualidade visual percebida pelo Sistema Visual Humano, HVS. A avaliação da qualidade de imagem é uma ferramenta necessária para fornecer um nível de qualidade com o objetivo de dar suporte a sistemas de processamento de imagem para melhoria, restauração, remoção de ruídos dos conteúdos visuais entregues aos usuários finais (OPOZDA; SOCHAN, 2014).

As métricas de avaliação da qualidade de imagem digitais são classificadas em subjetivas e objetivas. As métricas subjetivas são realizadas pelo julgamento subjetivo da imagem observada pelo HVS, já as objetivas são realizadas por algoritmos automatizados que buscam representar a avaliação subjetiva.

2.1.1 Avaliação subjetiva

A avaliação subjetiva é modo mais confiável para obter a estimativa da qualidade das imagens digitais, pelo fato de que os olhos humanos usualmente são os receptores finais das imagens (ZHAI; MIN, 2020). A avaliação subjetiva pode ser realizada, conforme a disponibilidade da imagem de referência, de dois modos: duplo estímulo ou estímulo único (PATIL; PATIL, 2018). A avaliação de duplo estímulo é a metodologia de avaliação subjetiva na qual o avaliador tem a sua disposição a imagem de referência (a imagem em sua forma original) e a imagem a ser testada (avaliada). Na avaliação de único estímulo, o avaliador só tem disponível a imagem avaliada, ou seja, não tem acesso a imagem de referência (ITU, 2012). As opiniões subjetivas são mapeadas em escalas, por exemplo, escala de qualidade

Tabela 1 – Escala categórica para ser utilizada na avaliação subjetiva de imagens digitais.

1	2	3	4	5
Péssima Qualidade	Baixa Qualidade	Média Qualidade	Boa Qualidade	Excelente Qualidade

Fonte: Elaborada pelo autor

(ITU, 2012), escala de degradação (SÁNCHEZ-FERREIRA et al., 2019) e escala relativa. A Tabela 1 apresenta uma escala categórica para classificação subjetiva da qualidade de imagem definida na metodologia de avaliação do *International Telecommunication Union* (ITU) (ITU, 2012).

A avaliação subjetiva da qualidade de imagem envolve processos de configuração de ambiente de visualização das imagens, recrutamento de avaliadores, classificação de assuntos e processamento das avaliações subjetivas obtidas (ZHAI; MIN, 2020). Na avaliação subjetiva, as imagens avaliadas são apresentadas a um número de pessoas e seus julgamentos individuais são coletados. No processamento de julgamento subjetivo é extraída a média das pontuações das avaliações fornecidas por todos os participantes para cada imagem testada a fim de obter o nível de opinião média, *Mean Opinion Score* (MOS) (ITU, 2012; THUNG; RAVEENDRAN, 2009). Contudo, o processo de julgamento subjetivo é demorado e custoso (BOVIK, 2009).

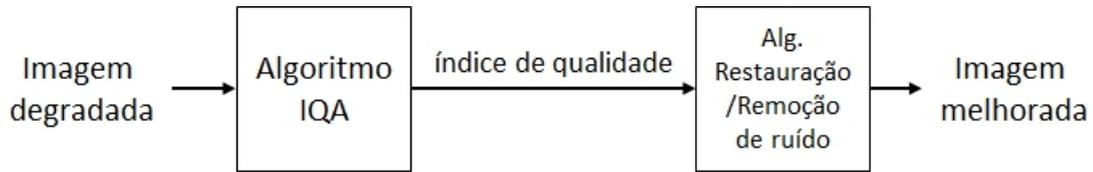
Os bancos de imagem de IQA e suas respectivas opiniões subjetivas, por exemplo, a média de opinião, são utilizados para treinar e testar o desempenho dos algoritmos objetivos. A avaliação das métricas objetivas de IQA em relação à opinião subjetiva é realizada por medidas de correlação, por exemplo, as correlações *Spearman Rank Order Correlation Coefficient* (SROCC) e o Coeficiente de Correlação Linear, *Linear Correlation Coefficient* (LCC) (CHEN; BOVIK, 2011).

2.1.2 Avaliação objetiva

Os métodos de avaliação da qualidade objetiva de imagens digitais são algoritmos que buscam representar a avaliação visual humana. A avaliação por um algoritmo permite indicar o nível de qualidade da imagem de forma automática e em menor tempo em comparação ao processo de avaliação subjetiva. A IQA objetiva é de ampla utilização em sistemas de processamento de imagens digitais (ZHAO et al., 2022; HU et al., 2022).

Um algoritmo de avaliação objetiva é útil para detectar o nível de degradação e dar

Figura 1 – Exemplo de aplicação que utiliza o índice de qualidade como uma etapa de um sistema de melhoramento da qualidade de imagem degradada.



Fonte: Elaborada pelo autor

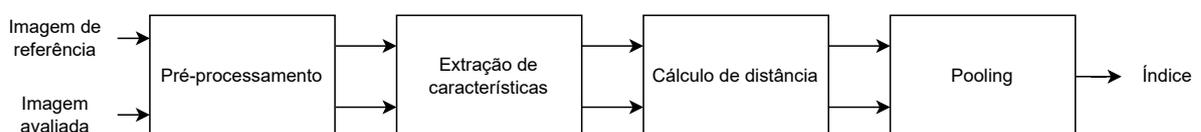
suporte, por exemplo, a algoritmos de remoção de ruídos ou de restauração de imagem (KAMBLE; BHURCHANDI, 2015). Um exemplo da utilização da avaliação da qualidade como suporte à restauração de imagens pode ser visto na Figura 1. O algoritmo IQA apresentado avalia a qualidade da imagem degradada e fornece um índice de qualidade. O índice de qualidade é utilizado como parâmetro de entrada para o algoritmo de restauração de imagem, que, por sua vez, executa seus passos para obter uma imagem melhorada.

2.1.3 Métodos de referência completa

Os métodos de avaliação de qualidade de imagem de referência completa (FR) realizam a avaliação da qualidade com acesso à informação da imagem de referência (PATIL; PATIL, 2018). Os métodos FR podem seguir as etapas: pré-processamento, extração de atributos, cálculo de distância, *pooling* e índice de qualidade. A extração de atributos pode ocorrer no domínio espacial, da transformada ou em ambos (ZHAI; MIN, 2020). A Figura 2 apresenta a estrutura geral de um *framework* utilizado pelos métodos de avaliação FR. As imagens passam por um pré-processamento e em seguida atributos, como luminância e contraste, são extraídos. Posteriormente, é realizado um cálculo de distância, por exemplo, a diferença, e, por fim, a etapa de agrupamento (*pooling*) é realizada.

Os primeiros métodos de avaliação de referência completa foram baseados no domínio espacial que consideram a imagem como um sinal bidimensional. A similaridade é cal-

Figura 2 – Estrutura de um *framework* utilizado por métodos de Avaliação da Qualidade de referência completa.



Fonte: Elaborada pelo autor

culada como uma medida de distorção (WANG; BOVIK, 2009). O erro médio quadrático (*Mean Squared Error*, MSE) é um dos métodos utilizados como medida de distorção entre duas imagens. Seja X a imagem de referência e Y a imagem avaliada. O MSE é definido por

$$MSE = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N [X(i, j) - Y(i, j)]^2, \quad (2.1)$$

em que M e N são as dimensões da imagem, i designa a i -ésima linha e j denota a j -ésima coluna de uma imagem (PREEDANAN et al., 2018).

A Razão Sinal-Ruído de Pico (*Peak Signal-to-Noise Ratio*, *PSNR*) é a medida que utiliza o MSE para avaliação da qualidade. O *PSNR* é definido como 10 vezes o logaritmo na base 10 da razão entre o quadrado do valor de pico (V_p) de amplitude do sinal de entrada e o erro médio quadrático,

$$PSNR (dB) = 10 \log_{10} \left[\frac{V_p^2}{MSE} \right]. \quad (2.2)$$

Assim, por exemplo, $V_p = 255$ é o nível máximo em tons de cinza de uma imagem 8bpp (bits/pixel) (WANG; BOVIK, 2009). O MSE indica o erro entre duas imagens, logo, espera-se um valor baixo para duas imagens semelhantes. O PSNR é inversamente proporcional ao MSE. Quanto maior o PSNR, espera-se que mais semelhantes sejam as imagens.

O cálculo do MSE é uma operação matemática simples e de baixa complexidade na sua execução. Um aspecto a se considerar, em se tratando do MSE e suas variações, é a baixa correlação que pode ser observada em relação à avaliação subjetiva. Este motivo se dá porque o MSE não leva em consideração os atributos percebidos pelo sistema visual humano (THUNG; RAVEENDRAN, 2009).

O alvo da avaliação objetiva é representar a avaliação subjetiva percebida pela visão humana como resultado da avaliação da qualidade, então, foi neste sentido que métricas de qualidade inspiradas no HVS combinaram atributos da imagem com algoritmos matemáticos (NIU et al., 2019b). Egiazarian *et al.* propuseram o PSNR-HVS que é a combinação do PSNR com as propriedades percebidas pelo sistemas visual humano (EGIAZARIAN et al., 2006). O PSNR-HVS considera a função de sensibilidade ao contraste, *Contrast Sensi-*

tivity Function (CSF). O HVS é menos sensível às distorções médias de brilho e variações de contraste em relação à outras distorções. Egiazarian *et al.* atribuíram modificações nos métodos PSNR-HVS e PSNR-HVSM para atribuir a característica da percepção visual, além do CSF, o mascaramento de contraste entre coeficientes da Discrete Cosine Transform (DCT), como resultado propuseram os métodos PSNR-HA e PSNR-HVS-HMA (PONOMARENKO *et al.*, 2011). Considere uma subimagem $g(i, j)$ de tamanho $M \times N$ cuja transformada discreta direta, $T(u, v)$, pode ser escrita em termos da relação geral

$$T(u, v) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} g(i, j)r(i, j, u, v), \quad (2.3)$$

para $u = 0, 1, 2, \dots, M - 1$ e $v = 0, 1, 2, \dots, N - 1$. $g(i, j)$ pode ser obtido utilizando a forma geral da transformada inversa discreta (GONZALEZ; WOODS, 2009)

$$g(i, j) = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} T(u, v)s(i, j, u, v), \quad (2.4)$$

para $i = 0, 1, 2, \dots, M - 1$ e $j = 0, 1, 2, \dots, N - 1$. A DCT é obtida substituindo os *kernels* (iguais) nas Equações 2.3 e 2.4,

$$r(i, j, u, v) = s(i, j, u, v) = \alpha(u)\alpha(v) \cos \left[\frac{(2i+1)u\pi}{2n} \right] \cos \left[\frac{(2j+1)v\pi}{2n} \right], \quad (2.5)$$

em que

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{n}} & \text{para } u = 0 \\ \sqrt{\frac{2}{n}} & \text{para } u = 1, 2, \dots, n - 1. \end{cases} \quad (2.6)$$

O sistema visual humano é sensível às informações estruturais da imagem. Partindo desta premissa, métodos de IQA extraem as estruturas da imagem e ao fim calculam a similaridade estrutural entre a imagem avaliada e de referência (ZHAI; MIN, 2020). O SSIM foi proposto para levar em consideração três canais de informações do HVS: contraste, brilho e estrutura (WANG *et al.*, 2004). Sejam X e Y a imagem original e a imagem distorcida respectivamente, o SSIM baseia-se no cálculo de três termos: luminância $l(X, Y)$, contraste $c(X, Y)$ e estrutura $s(X, Y)$ (WANG; BOVIK, 2009). O SSIM é definido pela combinação dos três termos:

$$SSIM(X, Y) = [l(X, Y)]^\alpha [c(X, Y)]^\beta [s(X, Y)]^\gamma. \quad (2.7)$$

A luminância, o contraste e a estrutura são dados respectivamente por

$$l(X, Y) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1}, \quad (2.8)$$

$$c(X, Y) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2}, \quad (2.9)$$

e

$$s(X, Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3}, \quad (2.10)$$

em que μ_X e μ_Y são as médias de X e Y , respectivamente, σ_X e σ_Y são os desvios padrões de X e de Y , respectivamente, e σ_{XY} é a covariância de X e Y , $\alpha = \beta = \gamma = 1$, e $C_3 = C_2/2$, $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$, $K_1 = 0,01$ e $K_2 = 0,03$. Em imagens em tons de cinza, L é o número de níveis de cinza da imagem (WANG; BOVIK, 2009).

Wang *et al.* afirmaram que a percepção dos detalhes da imagem dependem da densidade de amostragem do sinal, da distância do plano da imagem para o observador e da capacidade perceptiva do sistema visual do observador, desta forma, variações do SSIM foram propostas (WANG et al., 2004). Wang *et al.* sugeriram o *Multi-scale Structural Similarity* (MS-SSIM), que iterativamente aplica um filtro passa-baixa e realiza uma subamostragem (*downsampling*) nas imagens.

O domínio espacial do SSIM foi estendido para uma medida complexa de similaridade no domínio *wavelet* para formar o índice *Complex Wavelet SSIM* (CW-SSIM) (WANG; SIMONCELLI, 2005). O CW-SSIM foi desenvolvido para ser insensível às mudanças de luminância, contraste e transformações espaciais. Wang *et al.* propuseram *Combining Information Content Weighting SSIM* (IW-SSIM). Demonstraram que os pesos perceptivos ideais para agrupamento devem ser proporcionais ao conteúdo da informação local (ZHOUWANG2011; LI, 2011).

Zhang *et al.* (ZHANG et al., 2013) assumem que as informações semânticas da imagem estão fortemente representadas na borda de cada pixel e propuseram a métrica de qualidade *Edge Strength Similarity* (ESSIM). A distância de visualização da imagem para o avaliador e a resolução da imagem influenciam na avaliação da qualidade das imagens. Diante deste problema, Gu *et al.* introduziram uma etapa de pré-processamento e propuseram um novo modelo aplicado ao SSIM, o *Optimal Scale Selection SSIM* (OSS-SSIM) (GU et al., 2015).

Figura 3 – Modelos do gradiente Roberts.

-1	0	0	-1
0	1	1	0

Fonte: Elaborada pelo autor

O Structural Similarity envolve um alto custo computacional no termo do cálculo da luminância. Para reduzir a complexidade computacional, o Fast-SSIM foi proposto (CHEN; BOVIK, 2011). No Fast-SSIM, o termo de luminância utilizado pelo SSIM foi substituído utilizando uma janela de tamanho $n \times n$ para o cálculo da luminosidade entre a imagem original e a distorcida. O termo de luminância é extraído do valor médio de pixels dentro de uma janela quadrada. Seja $P(i, j)$ a soma dos valores de pixels acima e à esquerda do pixel (i, j) , incluindo o valor em (i, j) . O cálculo da soma sobre uma área do retângulo pode ser obtido com apenas duas adições e uma subtração, ou seja, $P(i, j) + P(i - m, j - n) - \{P(i, j - n) + P(i - m, j)\}$, em que m e n são a quantidade de pixels correspondentes à base e à altura do retângulo. A média dos pixels em $P(i, j)$ é definida por

$$\mu = \frac{P(i, j) + P(i - m, j - n) - \{P(i, j - n) + P(i - m, j)\}}{n^2}. \quad (2.11)$$

Outro consumo de tempo é o cálculo da variância e covariância. No algoritmo Fast-SSIM, os termos de variância e covariância foram substituídos pelo gradiente de Roberts (CHEN; BOVIK, 2011). A magnitude do gradiente é aproximada por

$$|\nabla f(x, y)| = \max\{|f(x + 1, y + 1) - f(x, y)|, |f(x + 1, y) - f(x, y + 1)|\} + \frac{1}{4} \min\{|f(x + 1, y + 1) - f(x, y)|, |f(x + 1, y) - f(x, y + 1)|\}, \quad (2.12)$$

em que $f(x + 1, y + 1) - f(x, y)$ e $f(x + 1, y) - f(x, y + 1)$ são magnitudes do gradiente Roberts nas duas direções ortogonais. A Figura 3 apresenta os modelos do gradiente Roberts.

A variância e a covariância entre duas imagens são

$$\sigma_{GX} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |\nabla X(i, j)| \quad (2.13)$$

e

$$\sigma_{GXY} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N |\nabla X(i, j)| |\nabla Y(i, j)|, \quad (2.14)$$

em que M e N são altura e largura da imagem, respectivamente. Os termos $|\nabla X|$ e $|\nabla Y|$ são os valores gradientes das imagens X e Y . O índice Fast-SSIM entre X e Y é definido por

$$\text{Fast-SSIM}(X, Y) = \frac{(2\mu_X\mu_Y + C_1)(2\sigma_{GXY} + C_2)}{(\mu_X^2 + \mu_Y^2 + C_1)(\sigma_{GX}^2 + \sigma_{GY}^2 + C_2)}. \quad (2.15)$$

O método *Most Apparent Distortion* (MAD) utiliza duas estratégias separadas para definir a avaliação da qualidade: a luminância e o contraste. A luminância local e o mascaramento de contraste são utilizados para estimar a distorção percebida baseada na detecção em imagens de alta qualidade e as alterações estatísticas locais dos componentes de frequência espacial são usadas para estimar as distorções baseadas na aparência em imagens de baixa qualidade (CHANDLER, 2010).

Em 2011, Zhang *et al.* afirmam que o HVS entende uma imagem de acordo com seus atributos de baixo nível. Partindo dessa premissa, foi proposto o FSIM (ZHANG; ZHANG; MOU, 2010). No FSIM a congruência de fase, *phase congruency* (PC), que é uma medida adimensional da significância de uma estrutura local, é utilizada como característica principal e a magnitude do gradiente, *gradient magnitude* (GM), é empregada como recurso secundário no FSIM. O valor do PC em um local pode refletir a probabilidade de ser um ponto de estrutura perceptivelmente significativo. Para um local x , se qualquer $f_1(x)$ e $f_2(x)$, onde f_1 e f_2 são imagens, tiver um valor PC significativo, implica que a posição x tem um alto impacto no HVS. O FSIM é definido como

$$FSIM = \frac{\sum_{x \in \Omega} S_L(x) \cdot PC_m(x)}{\sum_{x \in \Omega} PC_m(x)}, \quad (2.16)$$

em que Ω significa todo domínio espacial imagem, o $PC_m(x) = \max(PC_1(x), PC_2(x))$ pondera a importância de $S_L(x)$ entre as duas imagens. A similaridade $S_L(x)$ é definida por $S_L(x) = [S_{PC}(x)]^\alpha \cdot [S_G(x)]^\beta$, em que α e β são parâmetros usados para ajustar a importância relativa dos atributos PC e GM.

O sistema visual humano possui a capacidade de dar mais atenção a uma região saliente de uma cena antes de processá-la, ou seja, enfatiza-se um subconjunto de informações presente na imagem (ITTI; KOCH; NIEBUR, 1998). O *Visual Saliency-based Index* (VSI)

utiliza a saliência visual como característica no cálculo do mapa de qualidade local e emprega a saliência como uma função de ponderação para destacar a importância de uma região local (ZHANG; SHEN; LI, 2014).

Balanov *et al.* (BALANOV et al., 2015) propuseram uma métrica que avalia a qualidade de imagem medindo as alterações nas informações estruturais de sub-bandas no domínio da transformada discreta de cosseno. A *DCT Subbands Similarity* (DSS) pondera as estimativas de qualidade para essas sub-bandas. O *Mean Deviation Similarity Index* (MDSI) utiliza a similaridade de gradiente, similaridade de cromaticidade e desvio de *pooling* para medir mudanças na estrutura e distorções de cor (NAFCHI et al., 2016).

Sun *et al.* investigaram um conjunto de pixels chamado de super-pixel, que compartilham atributos visuais semelhantes e perceptivamente significativos para avaliação da qualidade. O *Superpixel-Based Similarity Index* (SPSIM) avalia a qualidade pela combinação dos atributos de similaridade de luminância de superpixel, similaridade da cromaticidade de superpixel e similaridade de gradiente de pixel (SUN et al., 2018).

Algoritmos de aprendizagem de máquina têm a capacidade de aprender e extrair atributos de imagens em diversos cenários aplicados ao processamento de imagem digital, como classificação, detecção de face, verificação de manuscritos, realce, super-resolução etc. Os métodos de avaliação de qualidade de imagem baseados em aprendizagem são compostos por duas fases: (i) pela coleta de atributos e (ii) pela regressão ou classificação utilizando estes atributos (NIU et al., 2019b).

O método *Support Vector Machine* (SVM) foi utilizado por (NARWARIA; Weisi Lin, 2012) para propor a métrica de Qualidade Baseada em Decomposição de Valor Singular, *Singular Value Decomposition* (SVD). O SVD divide a imagem em blocos e computada a singularidade para cada bloco. Os atributos formam um vetor de singularidade de blocos locais e globais que é aplicado ao SVM para obter o índice de qualidade das imagens.

Charrier *et al.* (CHARRIER; LÉZORAY; LEBRUN, 2012) propuseram a medida de qualidade baseada em aprendizagem de máquina *Machine Learning-based Image Quality Measure* (MLIQM) que utiliza um processo de combinação de classificadores baseado em SVM. Cada classificador corresponde a uma classe de qualidade ($[0, 1[$ - péssima, $[1, 2[$ - baixa, $[2, 3[$ - média, $[3, 4[$ - boa, $[4, 5]$ - excelente) (ITU, 2012). O método define uma função de regressão que projeta a pontuação da qualidade de uma imagem degradada dentro do intervalo associado $[0 - 5]$.

Liu *et al.* foram motivados pela observação de que não existe um único método capaz

de proporcionar o melhor desempenho em todos os tipos de distorções na avaliação da qualidade de imagem e propuseram a fusão de múltiplos métodos, *Multi-method Fusion* (MMF). O MMF é definido como a regressão linear do *score* ponderado a partir de outros métodos. Os índices são obtidos do processo de regressão (LIU et al., 2013).

Os gradientes das imagens são sensíveis à presença de distorções na imagem, assim como estruturas em diferentes locais de uma imagem distorcida sofrem diferentes graus de degradação (XUE et al., 2014). Partindo deste princípio, Xue *et al.* propuseram o método *Gradient Magnitude Similarity Deviation* (GMSD). O GMSD usa a estratégia de similaridade da magnitude do gradiente, *Gradient Magnitude Similarity* (GMS), pixel a pixel entre a imagem de referência e a imagem distorcida. O índice da avaliação da qualidade é obtido a partir do desvio padrão do mapa de GMS.

O método de aprendizagem *Extreme Learning Machine* (ELM) foi utilizado no método baseado na fatoração de matriz não negativa, *Non-negative Matrix Factorization* (NMF), para medir as degradações da imagem utilizando a representação baseada nas partes do NMF. O método aplica NMF nas imagens de referência e avaliada (distorcida) e os resultados das bases NMF são utilizados para formar vetores de similaridade de qualidade. O ELM é empregado para fundir os vetores de similaridade e avaliação subjetiva para prever a qualidade perceptiva final da imagem (WANG et al., 2017).

As redes neurais convolucionais, *Convolutional Neural Network* (CNN), têm a capacidade de extração de atributos a partir de n camadas de operações de convolução (GUO et al., 2020). As CNNs têm sido utilizadas para extração de atributos para avaliação objetiva de qualidade de imagem (BOSSE et al., 2018; KIM; LEE, 2017; ZHANG et al., 2018).

O Deep Image QuAlity Measure for FR IQA (DIQaM-FR) é um modelo CNN que foi proposto por (BOSSE et al., 2018). O DIQaM-FR é baseado na arquitetura VGGnet e possui dez camadas convolucionais, cinco camadas de *max-pooling* (2×2 pixel-size kernel) para extração de atributos e duas camadas completamente conectadas para regressão. As imagens de entrada são ajustadas em patches de tamanho 32×32 pixels.

O DIQaM-FR usa a estratégia de fusão para combinar os atributos extraídos da imagem de referência X (fx) e da imagem distorcida Y (fy). A fusão ocorre com a concatenação de atributos extraídos das imagens de entrada da rede, $concat(fx, fy)$. Outra estratégia é usar a diferença entre os atributos, $fy - fx$, $concat(fx, fy, fy - fx)$.

Np o número de *patches* selecionados da imagem e os recursos extraídos de cada *patch* são definidos como recursos visuais (v). O agrupamento por média simples assume que

todos os v têm a mesma importância na imagem para a estimativa global da qualidade. A qualidade global da imagem (q) é definida

$$q = \frac{1}{N_p} \sum_i^{N_p} v_i, \quad (2.17)$$

em que N_p denota o número de *patches* amostrados a partir da imagem e v_i é a i -ésima qualidade visual.

O *pooling* espacial obtido pela média das estimativas de qualidade local não considera o efeito da relevância perceptual espacialmente variável da qualidade local. Para considerar a variância espacial da qualidade relativa da imagem, foi proposto o agrupamento por Média Ponderada, *Image Quality Pooling by Weighted Average* (WaDIQaM). Nesta estratégia a rede emite uma ativação (a_i) para um *patch* i através da função de ativação ReLU e um termo de estabilidade $\epsilon = 1 \times 10^{-6}$,

$$a_i^* = \max(0, a_i) + \epsilon, \quad (2.18)$$

quando $a_i^* > 0$, é usado para ponderar a qualidade estimada a partir dos pesos normalizados,

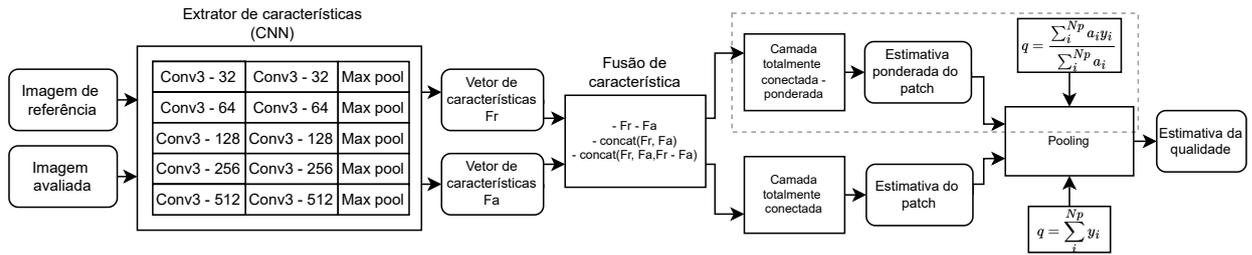
$$p_i = \frac{a_i^*}{\sum_j^{N_p} a_j^*}. \quad (2.19)$$

A qualidade global da imagem ponderada pode ser calculada por

$$q = \sum_i^{N_p} p_i v_i = \frac{\sum_i^{N_p} a_i^* v_i}{\sum_i^{N_p} a_i^*}. \quad (2.20)$$

A Figura 4 apresenta o modelo dos métodos DIQaM-FR e WaDIQaM-FR. O modelo recebe inicialmente as imagens de referência e avaliada e aplica 5 camadas convolucionais para extração de atributos. Os vetores de atributos extraídos da imagem de referência (Fx) e da imagem avaliada (Fy) são utilizados na fase de fusão de atributos. A fusão do atributos pode ser realizada com a simples diferença entre os vetores de atributos ($Fx - Fy$), ou pela concatenação (concat) dos vetores extraídos. Após a fusão, os vetores de atributos são apresentados a duas camadas totalmente conectadas. A estimativa do índice de qualidade utiliza duas estratégias de junção (*pooling*): simples média ou média

Figura 4 – Modelo dos métodos DIQaM-FR e WaDIQaM-FR (BOSSE et al., 2018).



Fonte: Elaborada pelo autor

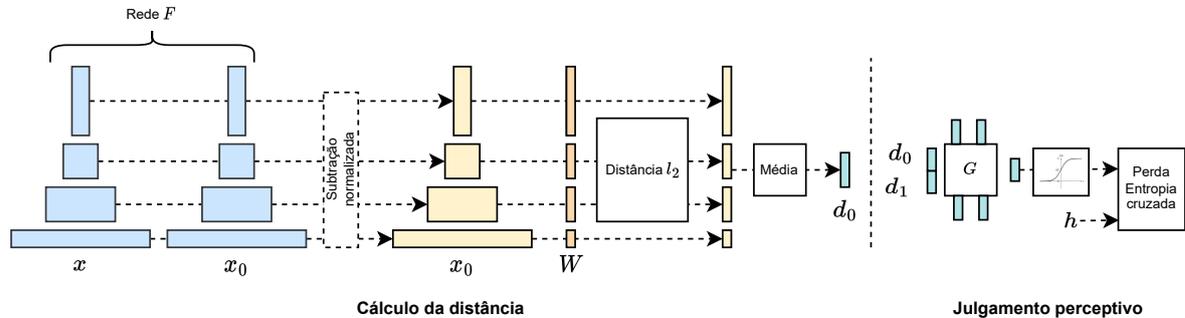
ponderada. A junção por média ponderada utiliza a multiplicação do vetor de característica pelo vetor de ativação pela Relu (a_i).

Deep Image Quality Assessment (DeepQA) foi proposto por (KIM; LEE, 2017), onde o comportamento do HVS é aprendido a partir da distribuição de dados subjacentes dos bancos de dados de IQA. O DeepQA tem como objetivo aprender as propriedades do sistema visual humano usando um mapa de erro que aloca pesos para os pixels de acordo com seus atributos espaciais locais. Para gerar um mapa de sensibilidade sem perder as informações de posição do pixel, a imagem distorcida e o mapa de erro passam por camadas convolucionais no início e depois são concatenados após a segunda camada convolucional. O DeepQA é composto por 6 camadas convolucionais e 2 camadas totalmente conectadas, as convoluções utilizam filtros 3×3 .

O *Learned Perceptual Image Patch Similarity* (LPIPS) foi proposto para alcançar a correspondência das preferências dos julgamentos perceptivos humanos a partir de atributos extraídos de redes de aprendizagem profunda (*deep learning*) (ZHANG et al., 2018). O LPIPS utiliza a distância de atributos (*features*) em diferentes arquiteturas de redes: SqueezeNet (IANDOLA et al., 2017), AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2017) e VGG. Especificamente neste trabalho foi utilizado a arquitetura VGG. A VGG utilizada foi composta por 5 camadas convolucionais.

A métrica LPIPS é composta por duas etapas: cálculo da distância entre blocos das imagens avaliadas e predição do julgamento perceptivo. A Figura 5 apresenta o diagrama de funcionamento da LPIPS. O cálculo da distância d é realizado entre os blocos das imagens de referência (x) e distorcida (x_0) com a rede F . Primeiramente as imagens x e x_0 são divididas em blocos de tamanho 32×32 . Posteriormente, é extraída uma pilha de recursos a partir de L camadas convolucionais e é normalizado cada unidade na dimensão do canal que é definido por $\hat{y}^l, \hat{y}_0^l \in \mathbb{R}^{M_l \times N_l \times C_l}$ para a camada l . As ativações por canal e

Figura 5 – Diagrama do funcionamento do método LPIPS composto por duas etapas: cálculo da distância e a previsão do julgamento perceptivo. Para calcular a distância d_0 entre dois blocos x, x_0 em uma rede convolucional F , primeiramente, os atributos da rede F são extraídos e são normalizados. As ativações por canal são escaladas pelo vetor W e calculada a distância l_2 . Por fim, a média espacial é calculada e somada por canal.



Fonte: Elaborada pelo autor

pelos vetores são escaladas, $w^l \in \mathbb{R}^{C_l}$, e calculada a distância l_2 . Finalmente, é calculada a média espacial e somados os resultados por canal. O cálculo da distância é definido por

$$d(x, x_0) = \sum_l \frac{1}{M_l N_l} \sum_{m,n} \|w_l \odot (\hat{y}_{mn}^l - \hat{y}_{0mn}^l)\|_2^2, \quad (2.21)$$

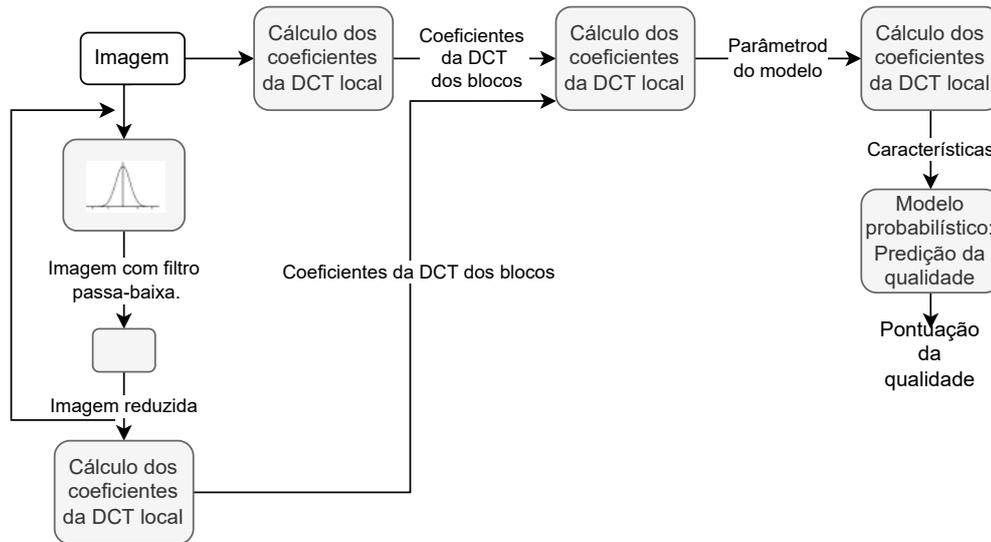
em que M e N corresponde a altura e largura da imagem.

2.1.4 Métodos não referenciados

Há cenários em que a imagem de referência não está disponível para a avaliação da qualidade. Métodos não referenciados, NR, de IQA fornecem a qualidade de uma imagem sem qualquer informação da imagem de referência (KAMBLE; BHURCHANDI, 2015). A IQA não referenciada é uma tarefa de maior complexidade em relação à avaliação *Full reference* (FR) e *reduced reference* RR.

Saad *et al.* propuseram o método não referenciado *Blind Image Integrity Notator Using DCT Statistics* (BLIINDS-II) (SAAD; BOVIK; CHARRIER, 2012). O BLIINDS-II é uma estrutura rápida de estágio único que se baseia no modelo estatístico de coeficientes DCT locais (SAAD; BOVIK; CHARRIER, 2012). A imagem avaliada é submetida ao cálculo dos coeficientes da DCT local. Essa etapa do pipeline consiste em particionar a imagem em blocos de tamanhos iguais $n \times n$, chamados de blocos da imagem local, e é calculada a DCT em cada bloco local. A extração do coeficiente é realizada localmente no domínio espacial

Figura 6 – Visão geral do framework Blinds-II.



Fonte: Elaborada pelo autor

baseado no fato de que o HVS processa o espaço visual localmente. A segunda etapa é aplicar um modelo de densidade gaussiana generalizada para cada bloco de coeficientes DCT. (SAAD; BOVIK; CHARRIER, 2011). A terceira etapa calcula as funções dos parâmetros do modelo derivado que são usados para prever as pontuações de qualidade da imagem. A última etapa usa um modelo *Bayesiano* simples para prever uma pontuação de qualidade. A Figura 6 apresenta uma visão geral do funcionamento das etapas da métrica BLINDS-II.

Mittal *et al.* propuseram um modelo de IQA não referenciado baseado em estatística de cena natural que opera no domínio espacial. O *Blind/Referenceless Image Spatial Quality Evaluator* (BRISQUE) não extrai atributos específicos de distorções, mas usa estatísticas de cena dos coeficientes de luminância normalizados localmente para quantificar as possíveis perdas de naturalidade (MITTAL; MOORTHY; BOVIK, 2011). O BRISQUE é baseado no princípio de que as imagens naturais possuem certas propriedades estatísticas regulares que são modificadas de forma mensurável pela presença de distorções. Primeiramente, o BRISQUE calcula a luminância normalizada localmente via subtração da média local e da normalização divisiva. Essa operação é aplicada a uma determinada imagem $Y(i, j)$ para produzir

$$\hat{Y}(i, j) = \frac{Y(i, j) - \mu(i, j)}{\sigma(i, j) + C}, \quad (2.22)$$

em que $i \in 1, 2, \dots, M$ e $j \in 1, 2, \dots, N$ são índices espaciais, M, N são a altura e largura da imagem, respectivamente, e C é uma constante que previne a instabilidade de ocorrência de divisão por zero. A média (μ) e desvio padrão (σ) são definidos como

$$\mu(i, j) = \sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} Y_{k,l}(i, j), \quad (2.23)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} (Y_{k,l}(i, j) - \mu(i, j))^2}, \quad (2.24)$$

em que $w = w_{k,l} = k - K, \dots, K, l = -L, \dots, L$ é uma função ponderada gaussiana circular simétrica $2D$ e $K = L = 3$. O pré-processamento na Equação 2.22 são definidos como coeficientes normalizados de contraste médio subtraído (MSCN). Mittal *et al.* afirmam que um distribuição gaussiana generalizada (GGD) pode ser usada para capturar um aspecto mais amplo de estatísticas de imagens distorcidas que geralmente exibem mudanças no comportamento da calda das distribuições de coeficientes. O GGD com média zero é dada por

$$f(y; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \left(- \left(\frac{|y|}{\beta} \right)^\alpha \right), \quad (2.25)$$

$$\beta = \sigma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}, \quad (2.26)$$

em que $\Gamma(\cdot)$ é uma função gama:

$$\Gamma(a) = \int_0^\infty t^{a-1} e^{-t} dt > 0. \quad (2.27)$$

Para cada imagem, o BRISQUE estima dois parâmetros (α, σ^2) a partir de um ajuste GGD dos coeficientes MSCN que formam o primeiro conjunto de atributos que serão usados para capturar a distorção da imagem. O BRISQUE modela as relações estatísticas entre os pixels vizinhos, no qual os coeficientes adjacentes exibem uma estrutura regular que é perturbada na presença de distorções. A estrutura de estatística de vizinhança foi modelada aos pares de coeficientes MSCN vizinhos em quatro orientações: horizontal (H),

vertical (V), diagonal principal (D1) e diagonal secundária (D2), conforme definido

$$\begin{aligned}
H(i, j) &= Y(i, j)I(i, j + 1) \\
V(i, j) &= Y(i, j)I(i + 1, j) \\
D1(i, j) &= Y(i, j)I(i + 1, j + 1) \\
D2(i, j) &= Y(i, j)I(i + 1, j - 1),
\end{aligned} \tag{2.28}$$

para $i \in 1, 2, \dots, M$ e $j \in 1, 2, \dots, N$. Além da GGD, o BRISQUE adota o modelo de Distribuição Gaussiana Generalizada Assimétrica (AGGD), definida por

$$f(y; \nu, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{\nu}{(\beta_l + \beta_r)\Gamma(\frac{1}{\nu})} \exp(-(\frac{-y}{\beta_l})^\nu) & x < 0 \\ \frac{\nu}{(\beta_l + \beta_r)\Gamma(\frac{1}{\nu})} \exp(-(\frac{y}{\beta_r})^\nu) & x \geq 0 \end{cases}, \tag{2.29}$$

em que

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}}, \tag{2.30}$$

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}}. \tag{2.31}$$

O parâmetro ν controla a forma de distribuição, enquanto σ_l^2, σ_r^2 são parâmetros escalares que controlam a dispersão, respectivamente. A média de distribuição η é definido por

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})}. \tag{2.32}$$

Assim, dezesseis parâmetros são calculados e utilizados em uma SVM para medir a qualidade da imagem (MITTAL; MOORTHY; BOVIK, 2011).

O *Codebook Representation for No-Reference Image Assessment* (CORNIA) é um método de NR-IQA de aprendizagem de característica que utiliza a estratégia não supervisionada (YE et al., 2012). O CORNIA constrói um *codebook* não rotulado de *patches* da imagem bruta por meio do algoritmo de agrupamento *k-means*. O *codebook* é utilizado na fase de codificação que utiliza um *soft-assignment* com o *max pooling*. Após o *pooling*, o CORNIA usa o regressor SVM para obter a índice de qualidade.

O método *Distortion Identification-based Image Verity and Integrity Evaluation* (DIIVINE) é baseado em uma estrutura de dois estágios envolvendo a avaliação da qualidade de imagem. Primeiramente, o DIIVINE identifica a distorção presente na imagem e, em seguida, realiza avaliação com base na distorção específica (MOORTHY; BOVIK, 2011). Os

atributos extraídos da imagem distorcida são usadas para classificar a imagem em n distorções. O conjunto de estatística é usado para avaliar a qualidade específica da distorção.

Mittal *et al.* propuseram o método *Naturalness Image Quality Evaluator* (NIQE) baseado na construção de uma coleção de atributos estatísticos “conscientes de qualidade” e ajustá-las em um modelo gaussiano multivariado, *MVG* (MITTAL; SOUNDARARAJAN; BOVIK, 2013). Os atributos de qualidade são derivados de um modelo estatístico de cena natural (NSS) simples do domínio espacial. O NIQE computa trinta e seis (36) atributos NSS de blocos do mesmo tamanho $n \times n$ da imagem a ser avaliada. Os atributos NSS a partir de blocos de imagens naturais podem ser obtidas ajustando-os com uma densidade MVG representado por

$$f_x(x_1, \dots, x_k) = \frac{1}{(2\pi)^{k/2} |\Sigma|^{1/2}} \times \exp\left(-\frac{1}{2}(x - \nu)^T \Sigma^{-1}(x - \nu)\right), \quad (2.33)$$

em que (x_1, x_2, \dots, x_k) são atributos calculados pelas Equações 2.25, 2.27, 2.29 e 2.32, ν e Σ denota a média e a matriz de covariância do modelo MVG. O tamanho do bloco pode variar de 32×32 até 160×160 pixels. A qualidade da imagem avaliada é definida por

$$D(\nu_1, \nu_2, \Sigma_1, \Sigma_2) = \sqrt{\left((\nu_1 - \nu_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (\nu_1 - \nu_2) \right)}, \quad (2.34)$$

em que ν_1, ν_2 e Σ_1, Σ_2 são vetores médios e matrizes de covariância do modelo MVG natural e do modelo MVG da imagem distorcida (MITTAL; SOUNDARARAJAN; BOVIK, 2013).

Liu *et al.* desenvolveram um método NR que utiliza atributos espaciais locais e entropia espectral para avaliação da qualidade. O *Spatial-spectral Entropy-based Quality* (SSEQ) tem uma estrutura baseada em dois estágios: (1) classificação da distorção e (2) um modelo de regressão. O SVM é usado para classificar e gerar a regressão dos atributos extraídos. O SSEQ divide a imagem em blocos de pixels e extrai *curvelet features* de cada bloco, produzindo um conjunto de vetores de atributos. Esses vetores são utilizados para calcular um vetor médio de atributos para produzir um vetor final e predizer a qualidade da imagem (LIU et al., 2014).

A métrica PIQUE é baseada no princípio de que a atenção da visão humana é direcionada diretamente para pontos salientes ou regiões espacialmente ativas da imagem. O PIQUE extrai atributos a partir da estatística de cena natural (NSS) segmentada em blocos não sobrepostos. Os blocos espacialmente ativos são classificados por dois tipos de

critérios de distorção: (1) critério de distorção perceptível e (2) critério de ruído branco. As pontuações de nível de bloco são agrupadas para determinar a qualidade da imagem (VENKATANATH et al., 2015). Os níveis de distorção são calculados em blocos locais de tamanho $n \times n$, onde $n = 16$. A Figura 7 apresenta o diagrama de blocos do funcionamento do método PIQUE. Inicialmente, a imagem é submetida a uma etapa de pré-processamento. Em seguida, o PIQUE realiza uma análise de nível de blocos para identificação de distorção. Os blocos identificados recebem uma pontuação com base no nível de distorção. As pontuações de nível de bloco são agrupadas para obter a qualidade da imagem avaliada. A etapa de pré-processamento é a remoção da média local e a normalização divisiva realizadas pelas Equações 2.23 e 2.24. Este passo extraí atributos NSS a partir da imagem Y em escala de cinza,

$$\hat{Y} = \frac{Y(i, j) - \mu(i, j)}{\sigma(i, j) + C}, \quad (2.35)$$

em que $i \in 1, 2, \dots, M$, $j \in 1, 2, \dots, N$ são índices espaciais, M e N são a altura e largura da imagem e $C = 1$ é uma constante para prevenir instabilidade. A imagem de entrada é segmentada em blocos não sobrepostos B_k , de tamanho $n \times n$, deixando de fora os blocos limites da imagem em todos os quatro lados da imagem. Os coeficientes MSCN são utilizados para rotular um determinado bloco como uniforme (U) ou espacialmente ativo (SA), ou seja, não uniforme. O critéria de rotulação dos blocos é dado a seguir

$$B_k = \begin{cases} U & \nu_k < T_U \\ SA & \nu_k \geq T_U. \end{cases}, \quad (2.36)$$

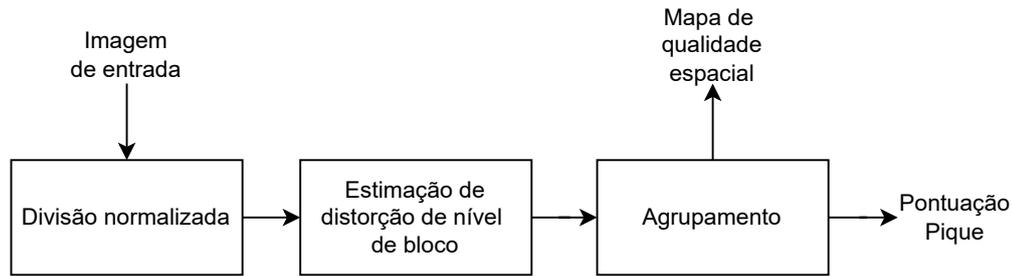
em que ν_k é a variação dos coeficientes MSCN, bloco B_k , $k \in 1, 2, \dots, N_B$, sendo N_B o total de blocos. A pontuação de qualidade do método PIQUE é definido por

$$PIQUE = \frac{\left(\sum_{k=1}^{N_{SA}} D_{sk} \right) + C_1}{(N_{SA} + C_1)}, \quad (2.37)$$

em que N_{SA} é o número de blocos espacialmente ativos em uma dada imagem e C_1 é uma constante positiva que exclui a possibilidade de divisão por zero. Os casos possíveis D_{sk} são

$$D_{sk} = \begin{cases} 1 & \text{se CDP e CR} \\ \nu_{blk} & \text{se CR} \\ (1 - \nu_{blk}) & \text{se CDP.} \end{cases}, \quad (2.38)$$

Figura 7 – Diagrama de blocos do método PIQUE.



Fonte: Elaborada pelo autor

em que CDP é o critério de distorção perceptível e CR é o critério de ruído.

O método *Integrated Local Naturalness Image Quality Evaluator* (IL-NIQE) integra atributos da estatística de imagens naturais de múltiplos sinais de entrada. Um modelo gaussiano multivalorado é treinado a partir de blocos de imagem extraídos de imagens naturais. Com o modelo gaussiano treinado, a distância de semelhança *Bhattacharyya* é utilizada para medir a qualidade de cada bloco da imagem e o índice de qualidade é obtido por um *pooling* médio (ZHANG; ZHANG; BOVIK, 2015).

Liu *et al.* propuseram o modelo *Oriented Gradients Image Quality Assessment* (OG-IQA), que é baseado em atributos de magnitude do gradiente. O OG-IQA utiliza a rede neural *Adaboosting* treinada com *back-propagation* para mapear os atributos em várias categorias de distorções. Assim, realiza-se a avaliação da qualidade da imagem (LIU *et al.*, 2016).

O método de uso geral *General-Purpose No-Reference Image Quality Assessment* (GP-NR-IQA) foi proposto como uma alternativa para os métodos de IQA voltados para distorções específicas. O GP-NR-IQA usa as estatísticas de planos de cores ortogonais, *Orthogonal Color Planes Pattern* (OCPP), para prever a avaliação da qualidade de propósito geral (FREITAS; AKAMINE; FARIAS, 2018a).

O método sem referência *Referenceless Image Quality Assessment* (RIQA), baseado no modelo de regressão *Gradient Boosting Machine* (GBM), foi proposto por Freitas *et al.* (FREITAS; AKAMINE; FARIAS, 2018b). O RIQA combina estatísticas do descritor de padrão de variação local de cor oposta (OC-LVP) com estatísticas do descritor de padrão saliente local de cor oposta (OC-LSP). O modelo GBM é usado na regressão do vetor de atributos obtidos pelo OC-LVP e OC-LSP e, assim, obtém o índice de qualidade da imagem.

O método não referenciado *Blind Image Blur Quality Evaluator* (BIBE) explora a

estatística da magnitude do gradiente da cena para modelar as propriedades de imagens desfocadas (*Blurred*). Primeiramente, o BIBE calcula o gradiente por um filtro *Prewitt* (3×3) e a distribuição de magnitudes do gradiente resultante é modelada por uma função gaussiana. Os parâmetros da função, como os de forma e escala, são usados para formar um vetor de atributos de 16 dimensões. O vetor de atributos é mapeado em índice de qualidade por um algoritmo de aprendizagem extrema, *Single-hidden Layer Feedforward Neural Networks* (SLFNs) (WANG et al., 2016).

Freitas *et al.* (FREITAS et al., 2020) se baseiam no fato de que as distorções visuais afetam as texturas da imagem e suas estatísticas. A partir desta suposição, Freitas *et al.* propuseram um método não referenciado de IQA baseado nos *Binarized Statistical Image Features* (BSIF), nos padrões binários locais completos, *Completed Local Binary Patterns* (CLBP), nos padrões de configuração local, *Local Configuration Patterns* (LCP) e na quantização de fase local, *Local Phase Quantization* (LPQ). O *Framework* usa a abordagem de aprendizado de máquina, *Machine learning* (ML), com treinamento supervisionado que leva em consideração as distorções que geram perturbação nas estatísticas pelos descritores BSIF, CLBP, LCP, e LPQ.

Niu *et al.* propuseram um modelo baseado em aprendizagem usando CNN siamesas, *Siamese Convolutional Neural Networks* (LRSN), para comparação da qualidade a partir de *patches* da imagem avaliada (NIU et al., 2019a). O modelo LRSN realiza uma normalização de contraste local na imagem. A imagem é dividida em *patches* do mesmo tamanho. Os *patches* são utilizados para treinar um modelo baseado em *Siamese Convolutional Neural Network* (SCNN).

Foi proposta uma abordagem de aprendizado multitarefa baseada em *deep learning* que consiste em duas tarefas: (1) tarefa de previsão da estatística de cenas naturais (NSS) e (2) tarefa de previsão da qualidade (YAN; BARE; TAN, 2019). Este modelo integra a tarefa de previsão de atributos de NSS à tarefa de previsão da qualidade baseada em *deep learning* para melhorar a habilidade de representação e generalização na avaliação da qualidade.

O método de Avaliação da Qualidade de Imagens não referenciado *Deep Image Quality Assessor* (DIQA) separa o treinamento em duas etapas: (1) uma parte objetiva, que foca nas distorções, e (2) uma parte relacionada ao sistema visual humano. Na primeira fase, a rede neural convolucional aprende a prever o mapa de erro objetivo, e, em seguida, o modelo aprende a prever a pontuação subjetiva (KIM; NGUYEN; LEE, 2019).

2.1.5 Avaliação da qualidade de imagem por amostragem

Na literatura, a estratégia de amostragem de imagem é utilizada em diversos problemas (MA; HAO, 2021; SHEN; KE, 2021; NEMA; PARSAL, 2022; HE et al., 2022). Na avaliação da qualidade em imagens digitais, a amostragem é realizada em duas estratégias: (1) pixel a pixel e (2) por bloco de pixel. A estratégia realizada pixel a pixel foi utilizada para avaliação da qualidade de imagem de referência completa para calcular o erro médio quadrático, MSE (LIMA; MADEIRO, 2009).

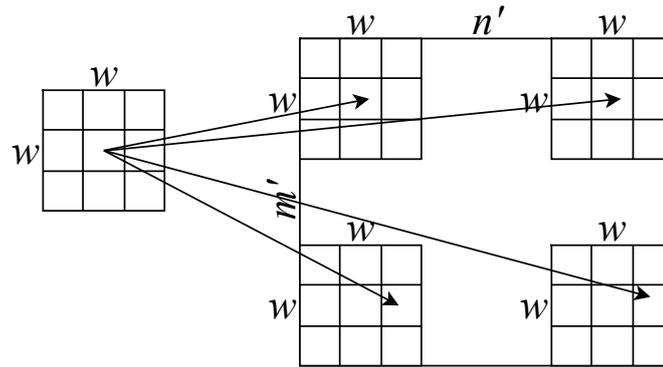
Um esquema de amostragem espacial aleatória de blocos de pixel foi utilizada para IQA de referência completa. O método utiliza uma amostragem de grade que substitui as operações de convolução por multiplicações locais baseadas em blocos (BAMPIS; GOODALL; BOVIK, 2016). Sejam X e Y imagens de referência e correspondente versão distorcida de $M \times N$ pixels. Um típico método de FR IQA, por exemplo SSIM, aplica um filtro $w \times w$ para obter as imagens filtradas X_f e Y_f . Para adequar o tamanho do filtro na imagem, foram definidos as dimensões \tilde{m} e \tilde{n} para a imagem, onde $\tilde{m} = M - w$ e $\tilde{n} = N - w$. No caso do SSIM, um subconjunto pode ser extraído usando amostradas aleatórias. Seja $N(p)$ o número de pixels selecionados,

$$N(p) = \left\lfloor \frac{p\tilde{m}\tilde{n}}{100} \right\rfloor, \quad (2.39)$$

em que $p \in (0, 100]$ é a porcentagem do tamanho da imagem ligeiramente cortada. Nesta amostragem, amostras podem ser selecionadas com ou sem repetição. Para obter uma avaliação de qualidade mais representativa, as amostras devem ser escolhidas de modo que não pertençam ao mesmo bloco de tamanho w . Para fazer isso, X e Y podem ser multiplicados diretamente por uma grade de filtros e, em seguida, adicionados e subtraídos adequadamente. Esta grade de filtros é criada repetindo o original filtro $w \times w$, onde $m' = \left\lfloor \frac{m}{w} \right\rfloor$, w vezes na direção vertical e $n' = \left\lfloor \frac{n}{w} \right\rfloor$, w vezes na direção horizontal. A Figura 8 apresenta um exemplo de grade de filtros.

Bruni *et al.* propuseram uma estratégia de amostragem baseada em entropia para acelerar a avaliação de imagens pelo SSIM (BRUNI; VITULANO, 2017). Bruni *et al.* definem que olhando para as imagens como fonte de informação é possível extrair um conjunto de blocos de pixels que represente as distorções visuais para ser aplicado ao SSIM. O processo de observação para descobrir blocos da imagem foi modelado como uma cadeia de Markov, o método caminha aleatoriamente sobre um grafo ponderado conectado cujos

Figura 8 – Exemplo de grade de filtros ($w \times w$) de tamanho $w = 3$.



Fonte: Elaborada pelo autor

nós são 2^l regiões R_1, R_2, \dots, R_{2^l} com pesos $W_{ij} \geq 0$ na aresta que une o nó i ao nó j . O algoritmo do método de Bruni *et al.* é apresentado abaixo:

- 1 Primeiramente, o método calcula a banda de aproximação wavelet A_G em G -ésimo nível da imagem X .
- 2 Aplica L níveis da Transformação de Quantização Média Sucessiva (SMTQ) para A_G e extrai as regiões R_1, R_2, \dots, R_{2^l} .
- 3 Calcula a cardinalidade n_1, n_2, \dots, n_l das regiões segmentadas e avalia os pesos do gráfico na equação

$$W_{ij} = \begin{cases} n_i & i = j \\ \frac{Z_{ij} + Z_{ji}}{2} & i \neq j, \end{cases} \quad (2.40)$$

em que $Z_{ij} = n_j \frac{\sum_{i \sim k, k \neq i} n_k}{\sum_{k=1}^{2^L} n_k}$ e n_i é o número de pixels da região R_i .

- 4 Extrai um ponto da região R_1 de acordo com a distribuição estacionária do gráfico como definido na Equação 2.40.
- 5 Calcula M_1 , ou seja, o SSIM sobre o bloco de dimensão $l \times l$ centralizado no ponto selecionado e atribua $k = 2$.
- 6 Extrai um ponto da região R_j selecionado de acordo com a probabilidade $P_{i,j} = \frac{W_{ij}}{\sum_{i \sim k} W_{ik}}$.
- 7 Calcula M_k , ou seja, o SSIM sobre o bloco de dimensão $l \times l$ centralizado no ponto selecionado.

8 Avalia o argumento

$$N_r = \operatorname{argmin}_k \frac{H(M_1, M_2, \dots, M_k)}{k} + \frac{k + 2\log_2(k) + 1}{2l^2}, \quad (2.41)$$

em que N_r é o comprimento de uma sequência no conjunto típico de distorção visual, l^2 é a dimensão do bloco e H é a entropia. Atribua o valor do argumento para a variável λ_k .

9 Se $\lambda_k > \lambda_{k-1}$, atribua $N_r = k - 1$ e $\widehat{M} = \frac{\sum_{k=1}^{N_r} M_k}{N_r}$, em que \widehat{M} é o valor aproximado do SSIM, e pare; caso contrário, atualiza $k = k + 1$ e vai para o passo 6.

3 SEQUÊNCIAS DE BAIXA DISCREPÂNCIA NA AMOSTRAGEM DE IMAGENS DIGITAIS

Neste capítulo são apresentadas as sequências de baixa discrepância Sobol e Van der Corput-Halton, bem como o *framework* para amostragem de imagens utilizando sequências uniforme, aleatórias e de baixa discrepância.

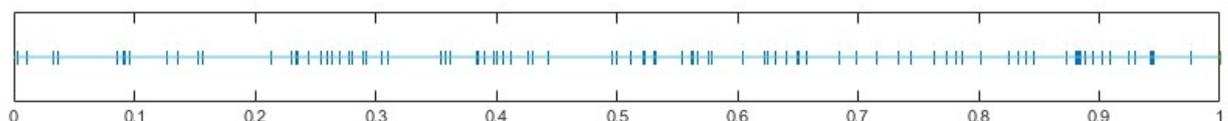
3.1 SEQUÊNCIAS ALEATÓRIAS

Sequências pseudo-aleatórias são sequências de números produzidos deterministicamente que exibem um comportamento aleatório quando vistos como um conjunto estatístico (WORLEY, 2016). O números pseudo-aleatórios, ou seja, as sequências aleatórias, são projetados em direção para ser mais "uniformemente" distribuídos que os números aleatórios puros (TEYTAUD, 2008).

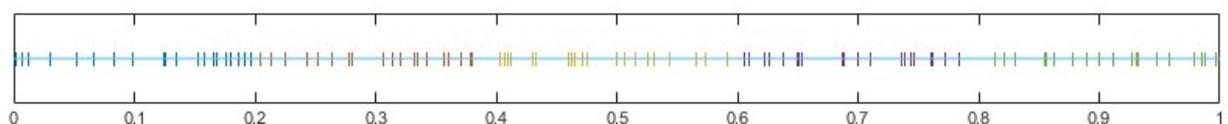
Números pseudo-aleatórios podem produzir pontos espaçados em uma região do intervalo, gerando lacunas entre os pontos, ou produzir pontos muito próximos, causando agrupamentos de pontos, em uma região específica do espaço. A Figura 9a apresenta 100 pontos aleatórios no intervalo $[0, 1]$. Como pode ser observado, há agrupamentos e espaços entre os pontos plotados.

Uma alternativa para melhorar a distribuição dos pontos aleatórios é dividir o espaço de amostragem em subespaços e gerar pontos sobre cada subespaço. A Figura 9b apresenta um exemplo de 100 pontos gerados em 5 subespaços do espaço. Foram plotados 20 pontos aleatórios gerados em cada subespaço. É possível observar uma melhor uniformidade de pontos nas cinco regiões destacadas em diferentes cores.

Figura 9 – 100 pontos aleatórios preenchendo todo o intervalo $[0, 1]$.



(a) Em todo o espaço.



(b) Em 5 subespaços dividindo o intervalo.

Fonte: Elaborada pelo autor

3.2 SEQUÊNCIAS DE BAIXA DISCREPÂNCIA

As sequências de baixa discrepância são geradas por uma família de algoritmos especiais para produzir pontos que são uniformemente distribuídos de acordo com uma medida de discrepância (CERVELLERA; MACCIO, 2013). A discrepância é um critério usado para medir a uniformidade de sequências no espaço d -dimensional (KRÖMER; PLATOŠ; SNÁŠEL, 2020). As sequências de baixa discrepância fornecem cobertura mais adequada dos pontos em uma região de interesse em comparação com sequências aleatórias (KUMAR; SUBRAMANIAN; SELVAM, 2010; LIMA; MADEIRO, 2009).

A medida de discrepância é definida como: Seja $\Omega = (\omega_1, \omega_2, \omega_3, \dots, \omega_n, \omega_{n+1}, \dots)$ uma sequência infinita de números reais no intervalo $[0, 1]$. Seja $I \subseteq [0, 1]$ um subintervalo do intervalo $[0, 1]$ (DALAL; STEFAN; HARWAYNE-GIDANSKY, 2008). Seja $A(I; n)$ a quantidade de pontos da subsequência $\omega_1, \omega_2, \dots, \omega_n$ pertencente a I , $A(I; n) = |I \cap \omega_1, \omega_2, \dots, \omega_n|$. Na sequência uniforme, a quantidade de pontos $A(I; n)$ seria equivalente a medida de I , ou seja,

$$\frac{A([\alpha, \beta]; n)}{n} = \beta - \alpha, \quad (3.1)$$

para $I = [\alpha, \beta)$ (CERVELLERA; MACCIO, 2013; LINS; LIMA; MELO, 2011). A discrepância dos primeiros n pontos da sequência Ω é definida como

$$D_n(\Omega) = \sup_{0 \leq \alpha \leq \beta \leq 1} \left| \frac{A([\alpha, \beta]; n)}{n} - (\beta - \alpha) \right|. \quad (3.2)$$

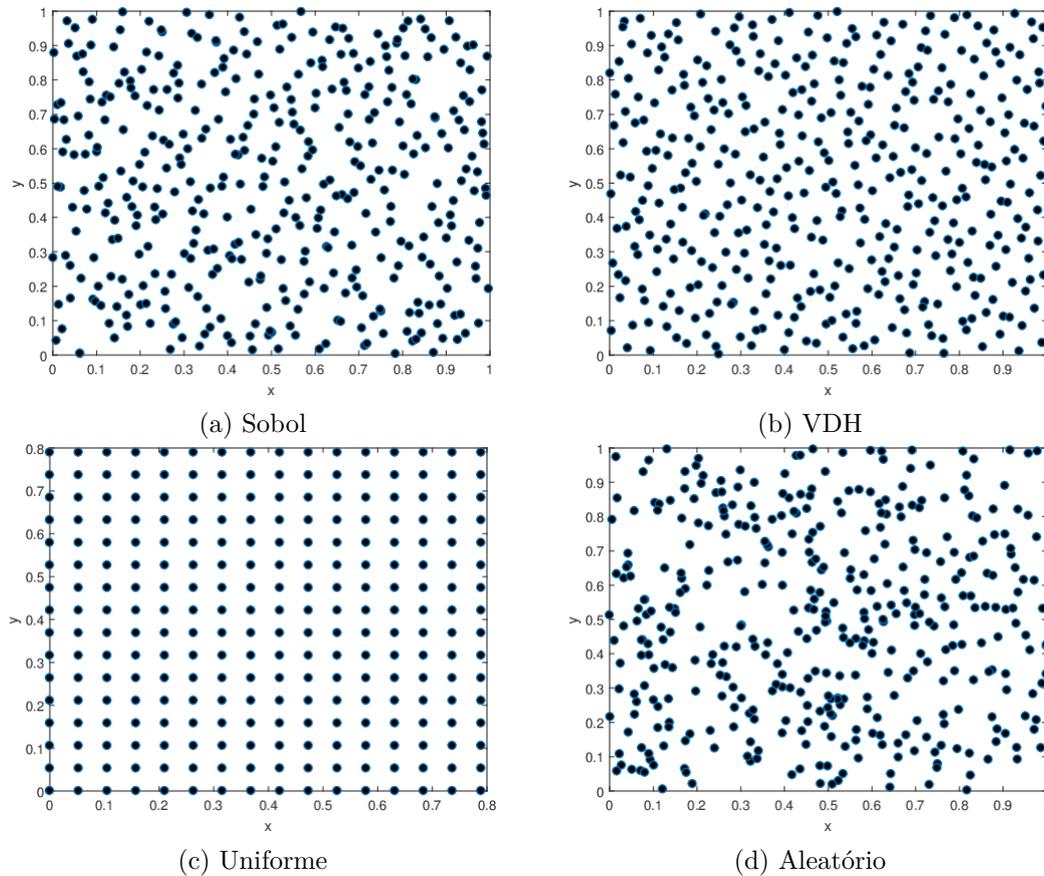
O cálculo da medida $D_n(\cdot)$ possui uma alta complexidade (LINS; LIMA; MELO, 2011). Utiliza-se uma medida mais simples, chamada discrepância-estrela, definida por

$$D_n^*(\Omega) = \sup_{0 < \beta \leq 1} \left| \frac{A([0, \beta]; n)}{n} - \beta \right|. \quad (3.3)$$

Uma sequência é de baixa discrepância quando tende a uma sequência uniformemente distribuída, ou seja, uma sequência infinita $\{\omega_n\}_{n=1}^{\infty} \subseteq I$ e de baixa discrepância quando $\lim_{n \rightarrow \infty} D_n = 0$ (LINS; LIMA; MELO, 2011). O objetivo da baixa discrepância é minimizar esta discrepância estrela (DALAL; STEFAN; HARWAYNE-GIDANSKY, 2008).

Diversas sequências de baixa discrepância são conhecidas mas algumas com um alto custo computacional para serem geradas. Neste trabalho, foram consideradas as sequências Sobol (SOBOL, 1967) e Van der Corput-Halton (VDH) (CORPUT, 1935) nas simulações, pelo baixo custo computacional na geração dos pontos de cada sequência.

Figura 10 – Gráficos de dispersão de 256 pontos bidimensionais no intervalo $[0, 1]$ obtidos das sequências Sobol, VDH, Uniforme e Aleatória. Os eixos horizontal e vertical são coordenadas dos pontos.



Fonte: Elaborada pelo autor

A Figura 10 apresenta gráficos de dispersão de 256 pontos bidimensionais no intervalo $[0, 1] \times [0, 1]$ obtidos das sequências Sobol, Van der Corput-Halton (VDH), pontos de distribuição uniforme e pontos pseudo-aleatórios. Neste exemplo, a distribuição aleatória gerou pontos espaçados ocasionando regiões sem cobertura de pontos e regiões com agrupamento dos pontos, consequentemente, os pontos aleatórios não garantem uma boa cobertura do espaço. Os pontos Sobol e VDH têm melhor distribuição e preenchimento na cobertura da área.

3.2.1 Sequência de Sobol

A sequência Sobol é uma sequência de baixa discrepância quase aleatória proposta por Sobol em 1967 (SOBOL, 1967). O algoritmo de sequência Sobol cria números pseudo-aleatórios a partir de polinômios primitivos $P(x)$ com coeficientes a_i a partir de dois elementos finitos (ou Galois) $GF(2)$, ou seja, binário (BRATLEY; FOX, 1988; DALAL; STE-

FAN; HARWAYNE-GIDANSKY, 2008). O j -ésimo ponto da sequência Sobol é definido a partir do polinômio inicial de grau s ,

$$P(x) = x^s + a_1x^{s-1} + a_2x^{s-2} + \cdots + a_{s-1}x + 1, \quad (3.4)$$

onde a_1, a_2, \dots, a_{s-1} são coeficientes com valores 0 ou 1, e

$$x^s = i_1q_1 \oplus i_2q_2 \oplus \cdots, \quad (3.5)$$

em que i e q_k são convertidos em binário, e \oplus denota uma operação de ou-exclusivo bit-a-bit. O i_k é o k -ésimo dígito a partir da direita $i = (i_k, \dots, i_4i_3i_2i_1)$. Os números de direção (q_1, q_2, \dots, q_k) são definidos por

$$q_k = \frac{M_k}{2^k}, \quad (3.6)$$

em que M_k é uma sequência de inteiros positivos $(M_1, M_2, M_3, \dots, M_k)$ a partir da relação:

$$M_k = 2a_1M_{k-1} \oplus 2^2a_2M_{k-2} \oplus \cdots \oplus 2^s a_1M_{k-s} \oplus M_{k-s}, \quad (3.7)$$

em que M_k , $1 \leq k \leq s$, é ímpar e menor que 2^k , e a_1, a_2, \dots, a_s são 0 ou 1 (BRATLEY; FOX, 1988).

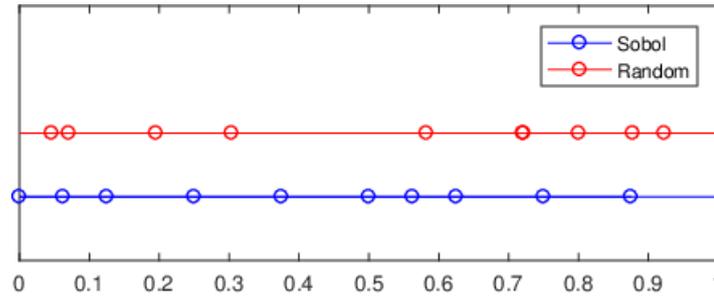
Um exemplo de sequência de pontos Sobol pode ser obtido atribuindo $s = 3$, $a_1 = 0$ e $a_2 = 1$ para formar o polinômio

$$p(x) = x^3 + x + 1, \quad (3.8)$$

que é usado para 1-dimensional pontos. Segundo o grau do polinômio, será $m_1 = 1$, $m_2 = 3$ e $m_3 = 7$. Os números de direção (q_k) são calculados

$$\begin{aligned} q_1 &= (0, 1)_2, \\ q_2 &= (0, 11)_2, \\ q_3 &= (0, 111)_2, \\ q_4 &= (0, 0101)_2, \\ q_5 &= (0, 00111)_2. \end{aligned} \quad (3.9)$$

Figura 11 – Exemplo de amostragem com 10 pontos gerados a partir das sequências Sobol e Aleatória.



Fonte: Elaborada pelo autor

Usando a Equação 3.7 e convertendo base binária para decimal temos

$$\begin{aligned}
 M_0 &= 0, \\
 M_1 &= (0, 1)_2 = 0,5_{10}, \\
 M_2 &= (0, 11)_2 = 0,75_{10}, \\
 M_3 &= (0, 1)_2 \oplus (0, 11)_2 = (0, 01)_2 = 0,25_{10}, \\
 M_4 &= (0, 111)_2 = 0,875_{10}, \\
 M_5 &= (0, 1)_2 \oplus (0, 111)_2 = (0, 011)_2 = 0,375_{10}.
 \end{aligned} \tag{3.10}$$

A Figura 11 ilustra os pontos gerados a partir de sequências pseudo-aleatórias e de baixa discrepância (Sobol). Neste exemplo, os pontos selecionados aleatoriamente formam grupos no intervalo $[0,0,1]$ e criam sobreposição $[0,7-0,75]$. Outra consequência dos pontos aleatoriamente é a presença de *gaps* entre os pontos $[0,3-0,6]$. A sequência Sobol, neste exemplo, preenche melhor a área de cobertura do espaço.

3.2.2 Sequência de Van der Corput-Halton

Van der Corput introduziu em 1935 (CORPUT, 1935) a construção de uma sequência no intervalo $[0, 1)$. Esta sequência é baseada na reflexão de dígitos binários de inteiros não negativos (FAURE; KRITZER; PILLICHSHAMMER, 2015). A sequência de Van der Corput foi proposta para gerar números em uma dimensão, Halton estendeu a sequência de Van der Corput para outras dimensões (HALTON, 1960). Esta sequência de números é distribuída no intervalo $[0,1]$ (DRMOTA, 2015). O VDH é definido como: Seja $b > 1$ um inteiro positivo que corresponde à base numérica. Seja A um número positivo decomposto na

Tabela 2 – Exemplo dos primeiros números da sequência VDH obtidos pela reversão (R) dos números da base binária.

$x_{(10)}$ (decimal)	$x_{(2)}$ (binário)	$Rx_{(2)}$ (binário)	$Rx_{(10)}$ (decimal)
0	0	0,0	0,0
1	1	0,1	0,5
2	10	0,01	0,25
3	11	0,11	0,75
4	100	0,001	0,125
5	101	0,101	0,625
6	110	0,011	0,375

Fonte: Elaborada pelo autor

base b , $A = a_0b^0 + a_1b^1 + \dots + a_jb^j$, em que a_j é a representação de A na base b . A sequência $VDH_b = (x_1, x_2, x_3, \dots)$ em que cada ponto x_i é escrito por

$$x_i = \sum_{j=0}^s \frac{a_j}{b^{j+1}}, \quad (3.11)$$

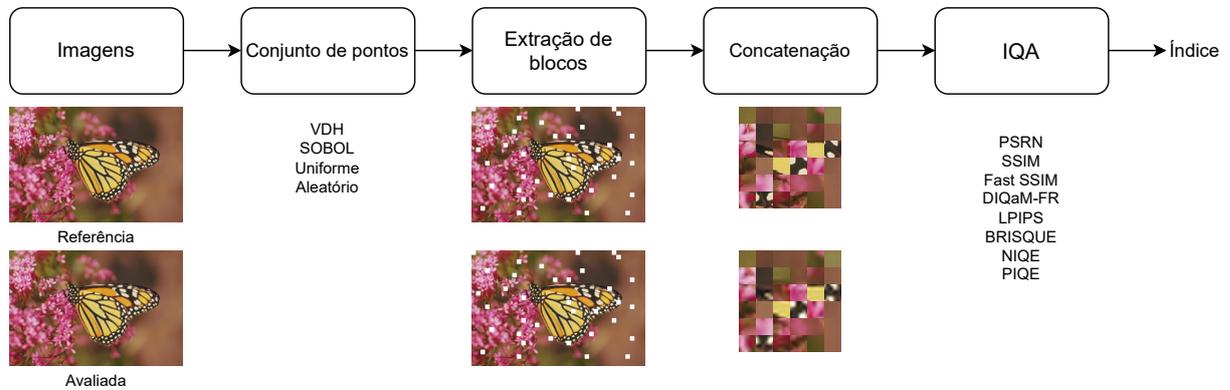
em que $\sum_{j=0}^s a_jb^j$ é a expansão na base b do número $i-1$. Um exemplo da sequência VDH para obter o número 6 na base 2 é apresentado. Primeiramente, o número 6 é convertido em binário 110, ou seja, $6 = 1 \times 2^2 + 1 \times 2^1 + 0 \times 2^0$. Para obter a sequência VDH para o número apresentado, invertamos cada fator e somamos 1 à potência que corresponde à base 2 do número 6 em binário

$$1 \times \frac{1}{2^{2+1}} + 1 \times \frac{1}{2^{1+1}} + 0 \times \frac{1}{2^{0+1}},$$

$$1 \times \frac{1}{8} + 1 \times \frac{1}{4} + 0 \times \frac{1}{2} = \frac{3}{8} = 0,375.$$

Uma outra forma que obter os números da sequência VDH é pela ordem reversa dos números na base binária. A Tabela 2 apresenta os primeiros números da sequência VDH obtidos a partir da reversão das casas dos números em sua forma binária.

Figura 12 – Estratégia de amostragem para avaliação da qualidade de imagens baseada em amostragem por blocos de pixel. A estratégia inicia com conjuntos de pontos que correspondem a um bloco na imagem. Em seguida, os blocos são extraídos, concatenados e apresentados à avaliação da qualidade.



Fonte: Elaborada pelo autor

3.3 MÉTODO DE AMOSTRAGEM DE IMAGENS COM SEQUÊNCIAS DE BAIXA DISCREPÂNCIA

O *framework* proposto para amostragem de imagens digitais utiliza sequências de pontos predefinidos, como as sequências de baixa discrepância, para coletar um conjunto de pixels espacialmente distribuídos na imagem com o objetivo de obter uma representação da imagem para um grupo menor de pixels. A representação por amostragem tem como objetivo produzir uma representação menor da imagem. As sequências de baixa discrepância Van der Corput-Halton (HDV) e Sobol foram selecionadas para fornecer um conjunto de pontos bidimensionais. Além das sequências de baixa discrepância, este trabalho também avaliou o desempenho das sequências uniforme e aleatória na amostragem de imagem.

As fases da estratégia de amostragem são: (1) obtenção de um conjunto de pontos, (2) extração de blocos de pixel, (3) concatenação dos blocos de pixel, e (4) avaliação da amostra por um método objetivo de IQA. A Figura 12 ilustra as fases da estratégia de amostragem proposta.

3.3.1 Conjuntos de pontos

A primeira fase da estratégia de amostragem é gerar uma sequência de pontos. As sequências Sobol, VDH, uniforme e aleatórias foram usadas para produzir os pontos de

Tabela 3 – Mapeamento de um ponto bidimensional para o espacial de uma imagem com dimensões de 512×384 pixels. As 1° e 2° dimensões do ponto são mapeadas para as dimensões da imagem.

Ponto bidimensional		Mapeamento espacial	
1°	2°	1°	2°
0,0	0,0	0	0
0,5	0,33	256	126
0,25	0,66	128	253
0,75	0,11	384	42
0,125	0,44	64	168

Fonte: Elaborada pelo autor

amostragem. Os pontos gerados são bidimensionais e cada dimensão se refere a uma coordenada da imagem. As sequências de baixa discrepância são geradas uma única vez e são mapeados para os diferentes tamanhos das imagens. Esta característica das sequências de baixa discrepância permite, além de uma boa cobertura da imagem, obter vantagem em comparação à geração de números aleatórios, visto que não há necessidade de gerar novos números a cada nova sequência pretendida.

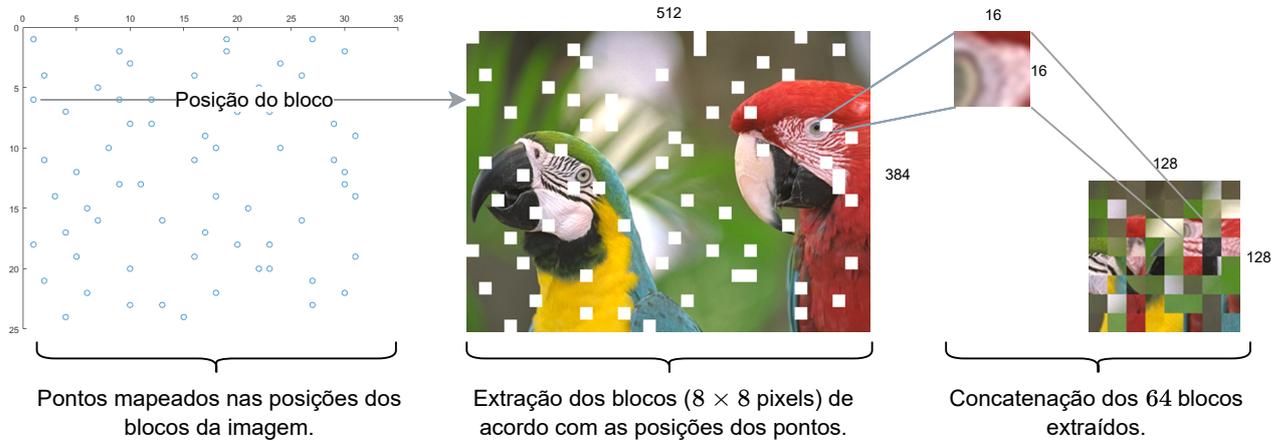
A Tabela 3 apresenta o mapeamento de cinco primeiros pontos da sequência VDH para o mapeamento espacial de uma imagem com dimensões de 512×384 pixels. As 1° e 2° dimensões do ponto são mapeadas para o comprimento e altura da imagem. O resultado do mapeamento é obtido da multiplicação do valor da dimensão do ponto pela dimensão da imagem.

3.3.2 Extração dos blocos de pixel

A fase de extração coleta os pixels correspondentes a cada ponto da sequência mapeado na imagem digital. Na amostragem pixel a pixel, cada ponto bidimensional corresponde a um pixel da imagem espacialmente localizada. Na amostragem de blocos de pixel, neste trabalho, foi definido que um ponto corresponde ao pixel no canto superior esquerdo do bloco de pixels a ser coletado. A representação da imagem com amostragem pode ser vista como um subconjunto da imagem. Sejam X uma imagem,

$$X = \{x_{(1,1)}, x_{(1,2)}, \dots, x_{(1,N)}, \dots, x_{(M,N)}\} \quad (3.12)$$

Figura 13 – Exmplo de extração de 64 blocos de tamanho de 16×16 pixels sobre a imagem I23 da base TID2013. Foram utilizados 64 pontos da sequência VDH mapeados sobre o espaço de 32×24 blocos de pixels da imagem.



Fonte: Elaborada pelo autor

em que $x_{(M,N)}$ é o último *pixel* da imagem, M, N são as dimensões da imagem X e P um conjunto de pontos bidimensionais não repetidos,

$$P = \{p_1, p_2, p_3, \dots, p_k\}, \quad (3.13)$$

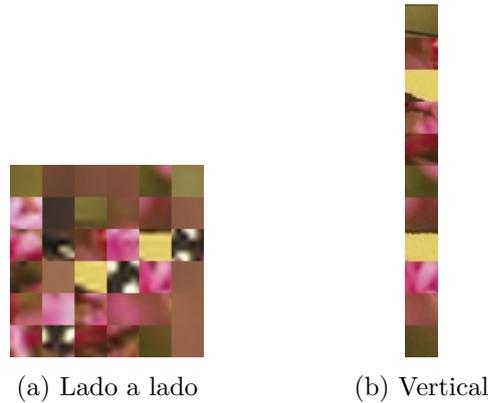
em que p_i , com $i = 1, 2, \dots, k$ é i -ésimo ponto bidimensional da sequência. X' é um subconjunto da imagem X que pode ser definido como

$$X' = \{x_{[p_{1(1)}, p_{1(2)}]}, x_{[p_{2(1)}, p_{2(2)}]}, x_{[p_{3(1)}, p_{3(2)}]}, \dots, x_{[p_{k(1)}, p_{k(2)}]}\}, \quad (3.14)$$

em que $p_{k(d)} | d \in (1, 2)$ é d -ésima dimensão do ponto p_k , para $p_{k(1)} < M$ e $p_{k(2)} < N$. A representação da imagem X é formada pelos pixels que formam a imagem X' .

A Figura 13 apresenta um exemplo de extração de blocos da imagem I23 da base TID2013. A imagem foi dividida em blocos de 16×16 e foram utilizados 64 pontos da sequência VDH para coletar 64 blocos de pixels da imagem. Os 64 pontos da sequência foram mapeados nas posições do espaço total de blocos da imagem (32×24 blocos de pixels). Cada bloco de pixels mapeado foi extraído e concatenado para formar um subconjunto da imagem, ou seja, uma imagem amostrada de tamanho de 128×128 pixels.

Figura 14 – Exemplos de concatenação lado a lado e vertical.



Fonte: Elaborada pelo autor

3.3.3 Concatenação dos blocos de pixel

Após a fase de extração, os blocos de pixel são concatenados formando a imagem amostrada, X' . A concatenação dos blocos de pixel é realizada posicionando os blocos à direita e abaixo (lado a lado) ou apenas uma concatenação vertical, ou seja, enfileirando os blocos de pixel sequencialmente.

Na concatenação lado a lado, os blocos de pixels podem ter 8 vizinhanças. No processo de extração de atributos, por exemplo, utilizando filtragem, pode ocorrer interferência de blocos que estavam localizado em regiões diferentes na imagem, ou seja, um bloco pode ser influenciado por pixels que estavam espacialmente distantes. Assim, é proposta a concatenação vertical para evitar a influência de blocos espacialmente distantes na extração de atributos da imagem amostrada. A estratégia de concatenação vertical permite no máximo 2 vizinhanças para cada bloco. A Figura 14 apresenta um exemplo de concatenação lado a lado e vertical.

3.3.4 Avaliação da qualidade das amostras

Após a fase de concatenação, as imagens amostradas são avaliadas e a qualidade é obtida por métodos de avaliação da qualidade objetivos. Os índices de qualidade obtidos a partir das imagens amostradas são avaliados para verificar se alcançam o mesmo nível de qualidade dos índices quando obtidos a partir das imagens com todos os pixels. Neste trabalho, os métodos de IQA objetivos utilizados para avaliar a estratégia de amostragem proposta foram selecionados pelo critério de ampla utilização dos métodos na avaliação

da qualidade de imagens com ou sem referência no estado da arte.

Os métodos de IQA utilizados nas simulações para obter o índice de qualidade das imagens amostradas foram os de referência completa: *Peak Signal-to-Noise Ratio* (PSNR) (WANG; BOVIK, 2009), *Structural Similarity* (SSIM) (WANG et al., 2004), *Fast Structural Similarity* (Fast-SSIM) (CHEN; BOVIK, 2010), *Deep Image Quality Measure for Full Reference Image Quality Assessment* (DIQaM-FR) (BOSSE et al., 2018), *Learned Perceptual Image Patch Similarity* (LPIPS) (ZHANG et al., 2018), e os métodos sem referência: *Blind/Referenceless Image Spatial Quality Evaluator* (BRISQUE) (MITTAL; MOORTHY; BOVIK, 2011), *Naturalness Image Quality Evaluator* (NIQE) (MITTAL; SOUNDARARAJAN; BOVIK, 2013) e *Perception-based Image Quality Evaluator* (PIQUE) (CHAN; GOLDSMITH, 2000).

4 EXPERIMENTOS E RESULTADOS

Este capítulo descreve os experimentos e protocolos realizados para medir o desempenho da estratégia de amostragem a partir de sequências de baixa discrepância na avaliação da qualidade de imagens digitais.

4.1 PROTOCOLO DOS EXPERIMENTOS

Inicialmente, foram considerados dois cenários: as imagens foram amostradas pixel a pixel e por blocos de pixel, e posteriormente foi realizada uma comparação dos resultados dos índices de qualidade das imagens para ambos os cenários. Um ponto relevante para se avaliar é o comportamento do histograma das imagens amostradas em relação ao histograma da imagem com todos os pixels. Partindo desse ponto, foi realizada a comparação entre os histogramas de imagens amostradas com sequências de baixa discrepância e os histogramas das imagens com todos os pixels.

Os resultados da avaliação das imagens amostradas com sequências de baixa discrepância foram comparados com os resultados utilizando as sequências Uniforme e Aleatória, e também foram comparadas com as transformações de redimensionamento (*resize*) e recorte (*cropped*). As imagens com redimensionamento usam a interpolação bicúbica e os recortes foram realizados no centro das imagens. A Figura 15 apresenta um exemplo de transformações de redimensionamento e recorte da imagem I22 de tamanho 512×384 pixels da base TID2013 (PONOMARENKO et al., 2015).

Os métodos de avaliação da qualidade com e sem referência para analisar o desempenho das estratégias de amostragem foram escolhidos pela ampla investigação na literatura.

Figura 15 – Exemplo de transformações da imagem I22 de tamanho 512×384 pixels da base TID2013. O redimensionamento e o recorte ficaram com 175×175 pixels após as transformações.



(a) Imagem I22 original.



(b) redimensionada.



(c) Recorte no centro.

Fonte: Elaborada pelo autor

As avaliações no cenário que há disponibilidade da imagem de referência foram realizadas utilizando os métodos de IQA: PSNR, SSIM, Fast-SSIM, e os métodos baseado aprendizagem DIQaM-FR e o LPIPS. No cenário de IQA sem disponibilidade da imagem de referência, foram selecionados os métodos BRISQUE, NIQE, PIQUE. Um ponto a destacar é que parte dos métodos não foi projetado para trabalhar com blocos de pixels independentes. Por fim, é medido o desempenho de tempo da estratégia de amostragem com sequências de baixa discrepância em relação ao tempo gasto de avaliação das imagem com todos os pixels.

O teste *Wilcoxon rank-sum* (LARSON; FARBER, 2015) foi utilizado para determinar se os valores de qualidade dos métodos IQA das imagens originais e amostradas são significativamente diferentes. A hipótese nula H_0 representa a afirmação de “não haver diferença entre os índices de qualidade” obtido dos métodos usando imagem completa e os índices obtidos pela avaliação das imagens com a estratégia de amostragem. Um nível significância de 5%, se $p\text{-valor} < 0,05$, então significa rejeitar a hipótese nula e aceitar a hipótese alternativa H_1 . Se o valor $p > 0,05$ significa que H_0 não pode ser rejeitado.

As correlações de *Spearman* (SROCC) e linear (LCC) foram utilizadas para comparar o grau de correlação entre as imagens amostradas e suas opiniões subjetivas. A correlação SROCC é definida como uma medida do nível de relação entre duas variáveis. O SROCC é calculado usando as entradas de amostras de dados emparelhadas. A fórmula é definida como

$$SROCC = 1 - \frac{6 \sum d^2}{n(n^2 - 1)}, \quad (4.1)$$

em que n é o número de entradas de dados emparelhados e d é a diferença entre os postos de uma entrada emparelhada. A correlação linear (LCC) é uma medida da direção de uma relação linear entre duas variáveis. O LCC é definido como

$$LCC = \frac{n \sum xy - (\sum x)(\sum y)}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}}, \quad (4.2)$$

em que n é o número de pares de dados (LARSON; FARBER, 2015).

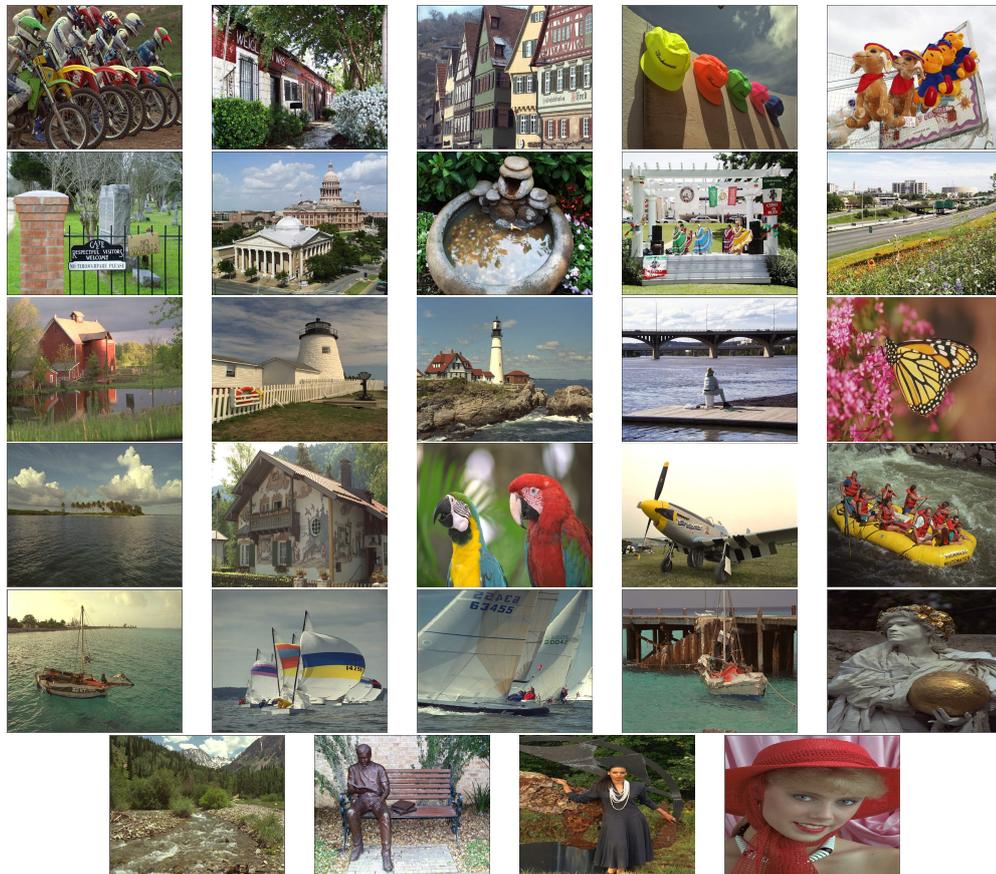
Os experimentos foram executados em um processador Intel Core I7 2.60GHz com 16GB de RAM e uma placa de vídeo de 6GB RTX2060. As simulações das métricas PSNR, SSIM, Fast-SSIM, BRISQUE, NIQE e PIQUE foram realizadas utilizando a ferramenta MATLAB. As métricas baseada em aprendizagem, DIQaM-FR, WaDIQaM-FR e LPIPS, utilizaram a linguagem PYTHON e executados com auxílio da placa de vídeo.

4.2 BANCOS DE IMAGENS

Os bancos de imagens utilizados para análise da estratégia de amostragem foram Live (HEIKH et al., 2018) e TID2013 (PONOMARENKO et al., 2015). Estes bancos de imagens são referências na análise de desempenho de métodos de IQA (CONG et al., 2022; CHEN et al., 2022; XU et al., 2022). As bases Live e TID2013 foram escolhidas pela ampla utilização na avaliação de métodos de avaliação da qualidade de imagens (BOSSE et al., 2018; WANG et al., 2004).

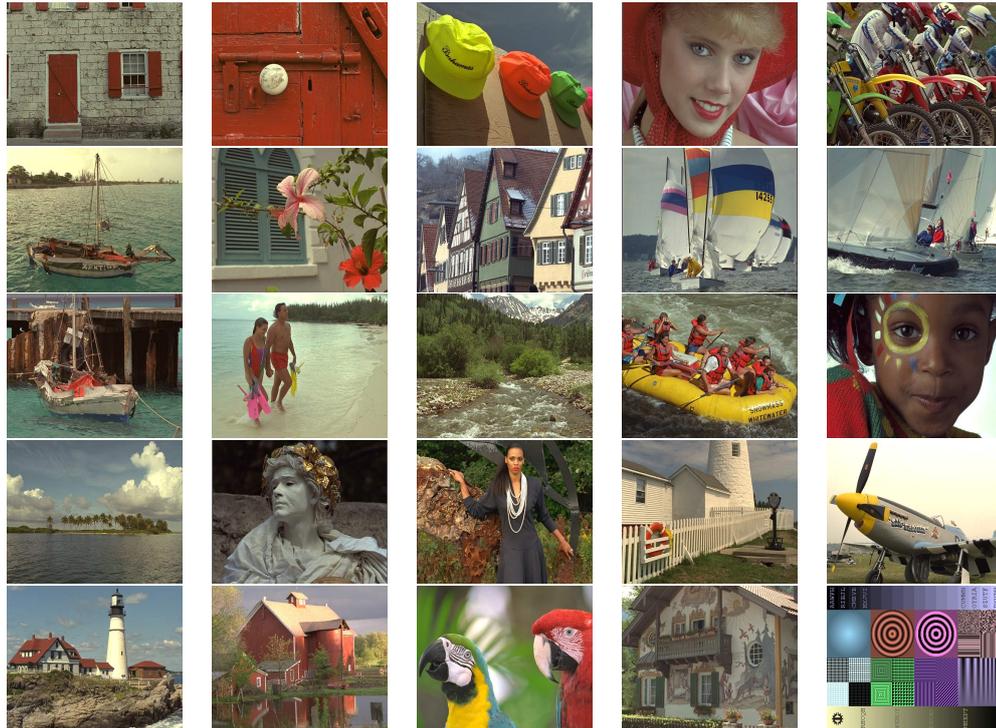
O banco de imagens Live inclui 982 imagens com distorções compressão JPEG2000 (227 imagens), JPEG (233 imagens), *white noise* (174 imagens), *Gaussian blur* (174 imagens) e *Fast-fading* (174 imagens) com *Difference Mean Opinion Score* (DMOS) (HEIKH et al., 2018). Os detalhes de como as imagens foram geradas com suas distorções podem ser encontrados em (SHEIKH; SABIR; BOVIK, 2006). As imagens que compõe a base Live têm dimensões 768×512 , 640×512 , 634×505 , 634×438 , 632×505 , 627×482 , 618×453 , 610×488 , 480×720 pixels. A Figura 16 apresenta as imagens de referência da base Live.

Figura 16 – Imagens de referência utilizadas nas simulações da base de imagens Live.



Fonte: (HEIKH et al., 2018)

Figura 17 – Imagens de referência utilizadas nas simulações da base de imagens TID2013.



Fonte: (PONOMARENKO et al., 2013)

O banco de imagens TID2013 (PONOMARENKO et al., 2013) é composto por 25 imagens de referência e 3000 imagens distorcidas. Para cada imagem de referência são aplicados 24 tipos de distorções e cada distorção com 5 níveis de intensidade. As distorções que compõe a base são Ruído gaussiano aditivo, ruído aditivo nos componentes de cor, ruído correlacionado espacialmente, ruído mascarado, ruído de alta frequência, ruído de impulso, ruído de quantização, desfoque Gaussiano, redução de ruído de imagem, compressão JPEG, compressão JPEG2000, erros de transmissão JPEG, erros de transmissão JPEG2000, ruído de padrão não excêntrico, distorções locais em bloco de diferentes intensidades, mudança média (mudança de intensidade), mudança de contraste, alteração da saturação da cor, ruído gaussiano multiplicativo, compressão com perdas de imagens com ruído, quantização de cores de imagem com *dither*, aberrações cromáticas, amostragem esparsa e reconstrução. Os detalhes de como as imagens foram geradas com suas distorções podem ser encontrados em (PONOMARENKO et al., 2015). As imagens que compõe a base têm dimensão de 512×384 pixels. A Figura 17 apresenta as imagens de referência utilizadas nas simulações da base TID2013.

4.3 AMOSTRAGEM PIXEL A PIXEL

O desempenho da amostragem pixel a pixel foi avaliado com a coleta de pixels utilizando os pontos obtidos das sequências de baixa discrepância VDH e Sobol e das sequências Uniforme e Aleatória. Os valores de SSIM das imagens amostradas com sequências de baixa discrepância foram comparados com os valores de SSIM obtidos das imagem com as transformações *Resize* e *Cropped*.

Na estratégia de amostragem pixel a pixel, foi definido um percentual de no máximo de 15% dos pixels das imagens avaliadas, pelo fato de que a transformação *Resize* e *Cropped* com a quantidade de pixels equivalente a taxas maiores de amostragem tende a se aproximar da medida verdadeira. As taxas de amostragem utilizadas nas simulações foram de 10×10 , 20×20 , 50×50 , 100×100 e 170×170 pixels. A base de imagens Live (HEIKH et al., 2018) foi selecionada para ser avaliada pelo método SSIM a fim de analisar o desempenho da amostragem pixel a pixel. Foram coletados as correlações SROCC, LCC, a média (μ) e o desvio padrão (σ) dos índices obtidos do SSIM da amostragem pixel a pixel. A Tabela 5 apresenta os valores obtidos da análise da amostragem pixel a pixel.

Os resultados obtidos com a estratégia de amostragem pixel a pixel revelam que as correlações SROCC e LCC foram distantes das correlações obtidas das imagens com todos os pixels avaliadas pelo SSIM. A amostragem Uniforme obtém melhor correlação SROCC em relação às estratégias de amostragem e os valores de média (μ) e desvio padrão (σ) destoam das médias e desvio padrão obtidos quando utilizado todos os pixels das imagens.

Na amostragem pixel a pixel com a estratégia de coleta de pixels com pontos Uniforme, à medida que se aumenta a quantidade de pixels, a representação da imagem tende a se aproximar a imagem original, e, conseqüentemente, a correlação tende a se aproximar da correlação obtida na avaliação das imagens com todos os pixels.

A Tabela 4 apresenta o desempenho das estratégias de amostragem das imagens avaliadas pelo SSIM com amostragem pixel a pixel na base Live com um total de 982 imagens. A amostragem Aleatória com uma taxa de 170×170 pixels alcançou apenas a correspondência de 16,70% (164 valores) comparado com os valores de SSIM quando avaliando as imagens com todos os pixel. Os valores da amostragem pixel a pixel Aleatória correspondem à média de 10 execuções, ou seja, as imagens são amostradas 10 vezes e calculado a média de correspondência dos valores do índice SSIM.

O teste de Wilcoxon demonstra que as imagens com amostragem e com transformações

resize e *cropped* não rejeitaram a hipótese nula (H_0), ou seja, os valores de SSIM obtidos das imagens amostradas não passaram no teste de hipótese quando comparado aos valores de SSIM das imagens com todos os pixels. Os resultados da amostragem pixel a pixel demonstram que os índices de SSIM a partir da avaliação das imagens amostradas não obtêm os mesmos índices de SSIM quando avaliadas as imagens com todos os pixels.

Tabela 4 – Quantitativo de correspondência do valor de SSIM usando amostragem pixel a pixel em relação ao valor de SSIM avaliando as imagens da base Live (total de 982 imagens) com todos os pixels.

Métodos	Taxa de amostragem				
	10×10	20×20	50×50	100×100	170×170
VDH	104	141	125	133	129
Sobol	105	151	141	142	140
Uniforme	119	106	41	19	15
Resize	81	81	81	81	84
Cropped	105	109	115	111	116
Aleatório	97	148	156	161	164

Fonte: Elaborada pelo autor

Tabela 5 – Avaliação das correlações SROCC, LCC e RMSE da amostragem pixel a pixel utilizando o método Structural Similarity aplicado à base Live. O μ denota a média dos valores de SSIM das imagens avaliadas, σ denota o desvio padrão e p corresponde ao p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.

Métodos	Amostragem	SROCC	LCC	μ	σ	RMSE	p
SSIM(Baseline)	Completa	0,9479	0,9124	0,8266	0,2035	47,7546	–
SSIM(VDH)	10 × 10	<i>0,8766</i>	<i>0,8272</i>	<i>0,9164</i>	<i>0,1572</i>	<i>47,7117</i>	<i>0,0000</i>
SSIM(Sobol)	10 × 10	0,8879	<i>0,8400</i>	<i>0,9167</i>	<i>0,1626</i>	<i>47,7098</i>	<i>0,0000</i>
SSIM(Uniforme)	10 × 10	<i>0,8729</i>	0,8412	<i>0,9100</i>	<i>0,1572</i>	<i>47,7430</i>	<i>0,0000</i>
SSIM(Resize)	10 × 10	<i>0,8469</i>	<i>0,4454</i>	<i>0,9841</i>	<i>0,0842</i>	<i>47,5953</i>	<i>0,0000</i>
SSIM(Cropped)	10 × 10	<i>0,7880</i>	<i>0,7231</i>	<i>0,8037</i>	<i>0,2608</i>	<i>47,8008</i>	<i>0,0326</i>
SSIM(Aleatório)	10 × 10	<i>0,8536</i>	<i>0,8402</i>	<i>0,9248</i>	<i>0,1501</i>	<i>47,6872</i>	<i>0,0000</i>
SSIM(VDH)	20 × 20	<i>0,8679</i>	<i>0,8543</i>	<i>0,8976</i>	<i>0,1641</i>	<i>47,7178</i>	<i>0,0000</i>
SSIM(Sobol)	20 × 20	<i>0,8911</i>	0,8789	<i>0,8937</i>	<i>0,1650</i>	<i>47,7193</i>	<i>0,0000</i>
SSIM(Uniforme)	20 × 20	0,9125	<i>0,8629</i>	<i>0,8167</i>	<i>0,1984</i>	<i>47,7637</i>	<i>0,0015</i>
SSIM(Resize)	20 × 20	<i>0,8816</i>	<i>0,5374</i>	<i>0,9852</i>	<i>0,0683</i>	<i>47,5960</i>	<i>0,0000</i>
SSIM(Cropped)	20 × 20	<i>0,8312</i>	<i>0,7547</i>	<i>0,7864</i>	<i>0,2245</i>	<i>47,7948</i>	<i>0,0006</i>
SSIM(Aleatório)	20 × 20	<i>0,8873</i>	<i>0,8742</i>	<i>0,8728</i>	<i>0,1777</i>	<i>47,6868</i>	<i>0,0000</i>
SSIM(VDH)	50 × 50	<i>0,8635</i>	<i>0,8492</i>	<i>0,9031</i>	<i>0,1641</i>	<i>47,7263</i>	<i>0,0000</i>
SSIM(Sobol)	50 × 50	<i>0,8741</i>	<i>0,8607</i>	<i>0,8999</i>	<i>0,1638</i>	<i>47,7281</i>	<i>0,0000</i>
SSIM(Uniforme)	50 × 50	0,9231	<i>0,8439</i>	<i>0,7835</i>	<i>0,2119</i>	<i>47,7977</i>	<i>0,0000</i>
SSIM(Resize)	50 × 50	<i>0,8971</i>	<i>0,7350</i>	<i>0,9773</i>	<i>0,0705</i>	<i>47,6039</i>	<i>0,0000</i>
SSIM(Cropped)	50 × 50	<i>0,8688</i>	<i>0,8019</i>	<i>0,7960</i>	<i>0,2108</i>	<i>47,7853</i>	<i>0,0003</i>
SSIM(Aleatório)	50 × 50	<i>0,8981</i>	0,8848	<i>0,8736</i>	<i>0,1745</i>	<i>47,6871</i>	<i>0,0000</i>
SSIM (VDH)	100 × 100	<i>0,8720</i>	<i>0,8593</i>	<i>0,8972</i>	<i>0,1656</i>	<i>47,7494</i>	<i>0,0000</i>
SSIM (Sobol)	100 × 100	<i>0,8908</i>	<i>0,8781</i>	<i>0,8941</i>	<i>0,1663</i>	<i>47,7494</i>	<i>0,0000</i>
SSIM (Uniforme)	100 × 100	0,9208	<i>0,8197</i>	<i>0,7580</i>	<i>0,2213</i>	<i>47,8233</i>	<i>0,0000</i>
SSIM (Resize)	100 × 100	<i>0,9111</i>	<i>0,8509</i>	<i>0,9586</i>	<i>0,0945</i>	<i>47,6227</i>	<i>0,0000</i>
SSIM (Cropped)	100 × 100	<i>0,8944</i>	<i>0,8335</i>	<i>0,7981</i>	<i>0,2044</i>	<i>47,7831</i>	<i>0,0000</i>
SSIM (Aleatório)	100 × 100	<i>0,9018</i>	0,8886	<i>0,8739</i>	<i>0,1735</i>	<i>47,6876</i>	<i>0,0000</i>
SSIM (VDH)	170 × 170	<i>0,8623</i>	<i>0,8478</i>	<i>0,9027</i>	<i>0,1637</i>	<i>47,7842</i>	<i>0,0000</i>
SSIM (Sobol)	170 × 170	<i>0,8716</i>	<i>0,8590</i>	<i>0,8987</i>	<i>0,1645</i>	<i>47,7865</i>	<i>0,0000</i>
SSIM (Uniforme)	170 × 170	<i>0,9105</i>	<i>0,7971</i>	<i>0,7402</i>	<i>0,2265</i>	<i>47,8408</i>	<i>0,0000</i>
SSIM (Resize)	170 × 170	0,9311	0,9073	<i>0,9320</i>	<i>0,1226</i>	<i>47,6493</i>	<i>0,0000</i>
SSIM (Cropped)	170 × 170	<i>0,9139</i>	<i>0,8498</i>	<i>0,7962</i>	<i>0,2036</i>	<i>47,7850</i>	<i>0,0000</i>
SSIM (Aleatório)	170 × 170	<i>0,9039</i>	<i>0,8905</i>	<i>0,8739</i>	<i>0,1733</i>	<i>47,6872</i>	<i>0,0000</i>

Fonte: Elaborada pelo autor

4.4 AMOSTRAGEM EM MÉTODOS DE REFERÊNCIA COMPLETA

4.4.1 Estratégia de amostragem no método *Structural Similarity*

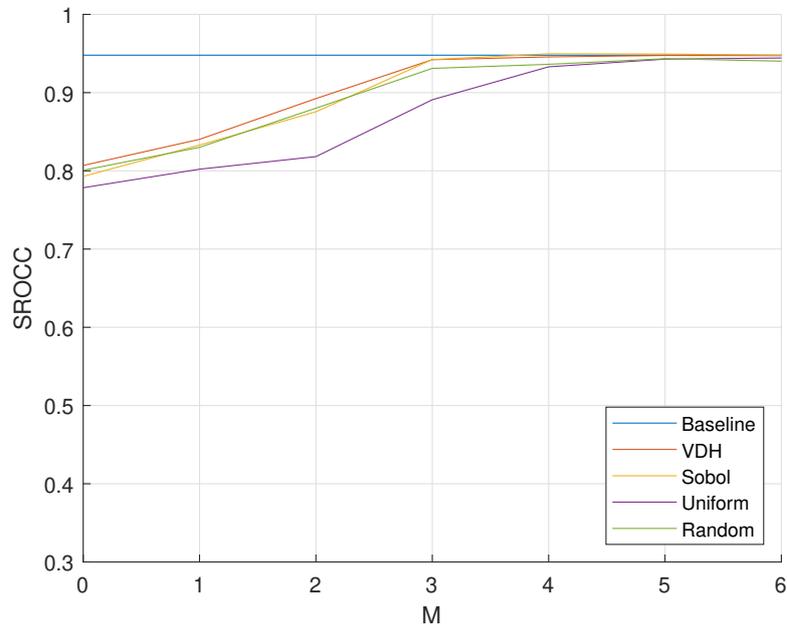
Na primeira análise, a estratégia de amostragem foi avaliada a partir do desempenho do algoritmo SSIM na avaliação da qualidade de imagens completa (todos os pixels) e com amostragem. Considerando a análise da amostragem pixel a pixel da Seção 4.3, foi definida a estratégia de amostragem por bloco de pixels. O tamanho do bloco para amostragem ocorreu pela investigação de diferentes tamanhos. Inicialmente, ao analisar a região de cobertura de pontos da imagem por uma sequência, foi observado que com 30 pontos possibilitava uma ampla cobertura da imagem. Foi definido um percentual de amostragem de 31,25% dos pixels da imagem que é equivalente a $30 \times 64 \times 64$ pixels de uma imagem de 512×768 pixels. Taxas maiores de amostragem levam à transformação *Resize* com o mesmo número de pixels a se aproximar da imagem original. Para investigar o melhor tamanho de bloco, foram fixados 30 pontos a serem coletados por blocos de pixels em cada imagem da base Live. O tamanho do bloco foi variado em $2^M \times 2^M | M = 0, 1, \dots, 6$ e as correlações SROCC e LCC foram obtidas.

As Figuras 18 e 19 apresentam o comportamento das correlações SROCC e LCC com o incremento do tamanho do bloco nas imagens amostradas da base Live e avaliadas pelo SSIM. As correlações SROCC e LCC a partir das imagens amostradas pelas sequências se aproximam dos valores reais com o tamanho do bloco de 8×8 ($M = 3$). É possível observar que com 30 blocos de 32×32 pixels as imagens com amostragem alcançam as mesmas correlações em comparação com as correlações SROCC e LCC usando todos os pixels das imagens.

Os valores médios dos índices obtidos na avaliação das imagens completas (Baseline) e na avaliação das imagens com amostragem pelo SSIM nos bancos de imagem Live e TID2013 foram investigados. Blocos de pixels foram utilizados para obter uma representação das imagens a partir das sequências Sobol, VDH, Uniforme e Aleatória para gerar as imagens amostradas. Os blocos foram concatenados de forma vertical. Os valores de SSIM dos blocos aleatórios foram obtidos da média de 20 execuções. Os valores de SSIM das imagens amostradas foram comparados com os valores de SSIM com transformações *Resize* e *Cropped*.

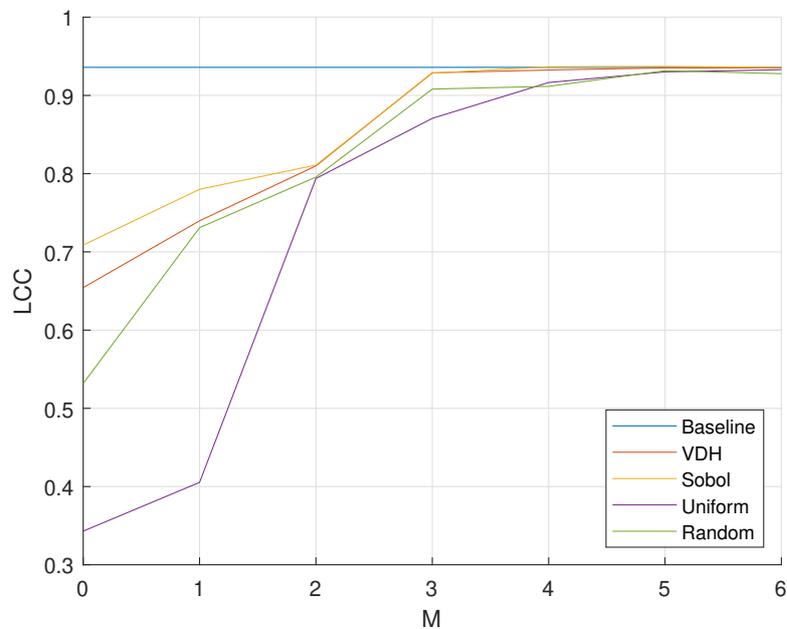
As imagens foram amostradas com 30 blocos de 32×32 pixels para o banco de dados

Figura 18 – Correlação SROCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatória variando o tamanho do bloco ($2^M \times 2^M$) na base de imagem Live. As imagens amostradas foram avaliadas pelo Structural Similarity.



Fonte: Elaborada pelo autor

Figura 19 – Correlação LCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatório variando o tamanho do bloco ($2^M \times 2^M$) na base de imagem Live. As imagens amostradas foram avaliadas pelo Structural Similarity.



Fonte: Elaborada pelo autor

Live e 12 blocos de 32×32 pixels para o banco de dados TID2013. As imagens que compõem a base TID2013 têm tamanho de 512×384 pixels e são menores que as imagens

que compõe a base Live, por isso, foi definido de forma empírica a partir de observações com diferentes quantidades de blocos. Para definir a quantidade de blocos coletados nas imagens da base TID2013, o tamanho do bloco foi fixado em 32×32 pixels, conforme experimento apresentado nas Figuras 18 e 19, e a quantidade de blocos foi variada de 5 – 30 blocos de pixels. Desta forma, foi definido que a quantidade de 12 blocos para a amostragem das imagens da base TID2013 alcançou resultados próximos quando avaliando as imagens com todos os pixels. O desempenho da estratégia de amostragem foi avaliado a partir da média (μ) e o desvio padrão (σ) dos índices obtidos do SSIM das imagens completas e amostradas. O p corresponde ao resultado do teste de Wilcoxon.

A Tabela 6 exhibe as correlações SROCC e LCC e os valores médios de SSIM das imagens com todos os pixels e das imagens amostradas a partir dos blocos de pixel dos bancos de imagens Live e TID2013. As imagens foram amostradas com 30 (7,80%) e 12 (6,25%) blocos de pixels do total de cada imagem das bases Live e TID2013, respectivamente. Os valores médios de SSIM (μ) demonstram que as imagens amostradas a partir de blocos de pixel apresentam valores próximos quando comparados com os valores de SSIM avaliando as imagens completa. Os valores destacados em itálico correspondem à rejeição da hipótese nula (H_0).

As médias dos valores da avaliação da qualidade pelo SSIM das imagens amostradas com as sequências VDH, Sobol e Uniforme foram próximas da média dos valores do SSIM quanto avaliando as imagens com todos os pixels (*Baseline*). A correlação SROCC entre o DMOS (*Difference Mean Opinion Score*) e os índices da avaliação das imagens com amostragem diferenciam da quarta casa decimal, ou seja, a correlação obtida a partir das imagens amostradas se aproxima da correlação obtida com todos os pixels das imagens, para as bases Live e TID2013. Os valores das correlações SROCC e LCC das imagens formadas com blocos aleatórios com 20 pontos de inicialização não atingem o mesmo desempenho para as bases Live e TID2013. O *Resize* foi a única estratégia que obteve rejeição da hipótese nula para o teste de Wilcoxon.

O teste estatístico de Wilcoxon foi aplicado para investigar se há diferenças entre os valores dos índices de qualidade obtidos a partir das imagens amostradas com as sequências VDH, Sobol, Uniforme, Aleatório e das transformações *Resize* e *Cropped*. A Tabela 7 apresenta os resultados obtidos da aplicação do teste de Wilcoxon sobre os valores dos índices obtidos. A aplicação do teste de Wilcoxon demonstrou que os índices obtidos das imagens amostradas com sequências Sobol, VDH, Uniforme, Aleatória e transformação

Cropped não obtiveram diferenças significativas. Os índices obtidos a partir das imagens com a transformação *Resize* diferem de todas as outras estratégias. Os valores dos índices obtidos a partir das imagens amostradas com sequência Aleatória foram obtidos pela média de 20 execuções.

O desempenho do SSIM foi comparado entre os índices obtidos das imagens completas e imagens amostradas. Foi computada a quantidade de valores de índices do SSIM em um intervalo de até 8% de diferença do valor original. Os valores do SSIM foram obtidos usando as bases Live e TID2013 com amostragem de blocos de pixels correspondendo a uma taxa de amostragem de 7,80% e 6,25%, respectivamente. As Tabelas 8 e 9 apresentam a quantidade de valores corretos de SSIM dentro do intervalo de até 8% de diferença do valor SSIM original (com todos os pixels). Nos bancos de dados Live (982 imagens) e TID2013 (3.000 imagens), o método de amostragem VDH alcançou 869 (88,49%) e 2583 (86,10%) de valores SSIM com uma diferença de apenas 2% em relação aos valores de SSIM originais. A estratégia de amostragem conseguiu a correspondência de 94,90% (932) de valores SSIM da Live e 95,93% (2878) da TID2013 com um erro de 5%.

4.4.2 Estratégia de amostragem no método *Peak Signal-to-Noise Ratio*

A avaliação do método PSNR foi realizada usando as mesmas taxas de amostragem e tamanhos de blocos utilizados na avaliação do método SSIM, sendo 30 blocos para a base de imagens Live e 12 blocos para a base de imagens TID2013. O tamanho dos blocos amostrados foram de 32×32 pixels. O tamanho das amostras corresponde aproximadamente

Tabela 6 – Valores médios de SSIM e correlações SROCC e LCC a partir das imagens amostradas usando sequências de baixa discrepância sobre as bases Live (30 blocos) e TID2013 (12 blocos). As imagens foram amostradas com blocos de tamanhos de 32×32 pixels. O μ denota a média dos valores do SSIM, σ denota o desvio padrão e p corresponde p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.

Algoritmo	Live					TID2013				
	SROCC	LCC	μ	σ	p	SROCC	LCC	μ	σ	p
SSIM (Baseline)	0,9479	0,9124	0,8624	0,1944	–	0,7417	0,7399	0,9039	0,1152	–
SSIM (VDH)	0,9470	0,9104	0,8657	0,1884	0,7533	0,7339	0,7338	0,9064	0,1113	0,1913
SSIM (Sobol)	0,9476	0,9114	0,8636	0,1905	0,8268	0,7413	0,7391	0,9055	0,1105	0,8431
SSIM (Uniforme)	0,9463	0,9065	0,8626	0,1921	0,7313	0,7302	0,7290	0,9044	0,1124	0,8308
SSIM (Resize)	<i>0,9313</i>	<i>0,9007</i>	<i>0,9530</i>	<i>0,1142</i>	<i>0,0000</i>	<i>0,7202</i>	<i>0,7693</i>	<i>0,9747</i>	<i>0,0489</i>	<i>0,0000</i>
SSIM (Cropped)	0,9390	0,8960	0,8537	0,2013	0,5167	0,7343	0,7318	0,8984	0,1283	0,1770
SSIM (Aleatório)	0,9470	0,9098	0,8673	0,1890	0,6102	0,7384	0,7363	0,9044	0,1222	0,0515

Fonte: Elaborada pelo autor

Tabela 7 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e avaliadas pelo método SSIM. O teste de Wilcoxon foi aplicado. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferença significativa entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com a estratégia de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,9669	0,3101	<i>0,0000</i>	0,4674	0,8007
Sobol	0,9669	0	0,3372	<i>0,0000</i>	0,4953	0,7571
Uniforme	0,3101	0,3372	0	<i>0,0000</i>	0,7934	0,2170
Resize	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	0	<i>0,0000</i>	<i>0,0000</i>
Cropped	0,4674	0,4953	0,7934	<i>0,0000</i>	0	0,3472
Aleatório	0,8007	0,7571	0,2170	<i>0,0000</i>	0,3472	0

Fonte: Elaborada pelo autor

Tabela 8 – O número de imagens com o valor SSIM dentro do intervalo de até 8% de diferença do valor SSIM original usando uma porcentagem de pixels das imagens na base Live, total de 982 imagens.

Método de amostragem	1%	2%	3%	4%	5%	8%
VDH	808	869	894	911	924	945
Sobol	784	849	904	920	932	950
Uniforme	732	810	856	886	905	933
Resize	363	405	445	478	515	595
Cropped	561	644	706	742	763	827
Aleatório	702	800	827	887	879	934

Fonte: Elaborada pelo autor

Tabela 9 – O número de imagens com o valor SSIM dentro do intervalo de até 8% de diferença do valor SSIM original usando uma porcentagem de pixels das imagens na base TID2013, total de 3000 imagens.

Método de amostragem	1%	2%	3%	4%	5%	8%
VDH	2305	2583	2746	2823	2878	2957
Sobol	2179	2485	2646	2758	2822	2927
Uniforme	2039	2356	2535	2663	2753	2886
Resize	1078	1347	1534	1698	1852	2114
Cropped	1678	2049	2235	2386	2485	2687
Aleatório	1633	1987	2250	2401	2341	2678

Fonte: Elaborada pelo autor

7,8% e 6,25% dos pixels das imagens para as bases Live e TID2013, respectivamente.

A Tabela 10 apresenta as correlações SROCC e LCC, média (μ) e desvio padrão (σ) do desempenho do PSNR nas bases Live e TID2013. As correlações SROCC e LCC entre os valores de qualidade obtidos pelo PSNR avaliando as imagens amostradas com as sequências VDH, Sobol e Uniforme são próximos aos valores de qualidade quando o PSNR avalia as imagens com todos os pixels. Os valores destacados em itálico correspondem à rejeição da hipótese nula (H_0). O resultados das correlações SROCC e LCC das imagens amostradas com blocos aleatórios são as médias dos resultados de 20 vezes os pontos de inicialização.

A amostragem Sobol obteve a maior correlação SROCC (0,8757) na base Live. A amostragem Uniforme obteve 0,4830 de correlação LCC na base TID2013. A média e o desvio padrão dos valores de qualidade obtidos pelo PSNR utilizando as sequências VDH, Sobol e Uniforme são próximos aos valores de referência de média e desvio padrão quando avaliado as imagens com todos os pixels. As imagens com a transformação *Resize* tem maior SROCC sobre a base Live do que as outras estratégias de amostragem. Entretanto, o teste de hipóteses demonstra que os índices obtidos da avaliação pelo PSNR das imagens redimensionadas (*Resize*) diferem dos índices obtido a partir da imagem completa. Na base TID2013, as sequências Uniforme, Aleatório e a transformação *Resize* foram rejeitados no teste de hipóteses ($p < 0,05$).

O teste Wilcoxon foi aplicado sobre a base Live para analisar se há diferenças significativas entre os valores dos índices obtidos a partir das sequências VDH, Sobol, Uniforme,

Tabela 10 – Valores médios de PSNR e correlações SROCC e LCC das imagens amostradas usando sequências de baixa discrepância sobre a base Live e TID2013. As imagens foram amostradas com 30 blocos para Live e 12 blocos para a TID2013, o tamanho do bloco foi de 32×32 pixels. O μ denota a média dos valores do PSNR, σ denota o desvio padrão e p corresponde p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.

Algoritmo	Live					TID2013				
	SROCC	LCC	μ	σ	p	SROCC	LCC	μ	σ	p
PSNR (Baseline)	0,8756	0,8585	27,4992	7,4504	–	0,6394	0,4785	30,0126	7,1790	–
PSNR (VDH)	0,8743	0,8583	27,5591	7,4896	0,9291	0,6317	0,4813	30,1226	7,4243	0,6741
PSNR (Sobol)	0,8757	0,8578	27,5805	7,4296	0,9724	0,6251	0,4722	29,9121	7,3355	0,4942
PSNR (Uniforme)	0,8702	0,8557	27,6456	7,5592	0,7782	<i>0,5552</i>	<i>0,4132</i>	<i>32,2686</i>	<i>8,8390</i>	<i>0,0000</i>
PSNR (Resize)	<i>0,9296</i>	<i>0,9223</i>	<i>36,3991</i>	<i>9,4653</i>	<i>0,0000</i>	<i>0,5851</i>	<i>0,5664</i>	<i>38,9625</i>	<i>9,3616</i>	<i>0,0000</i>
PSNR (Cropped)	0,8607	0,8471	27,1698	7,6459	0,2082	0,5962	0,4716	30,0006	7,8855	0,4741
PSNR (Aleatório)	0,8706	0,8547	27,7112	7,5082	0,3296	0,6194	0,4741	30,5253	7,6521	0,2082

Fonte: Elaborada pelo autor

Tabela 11 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e avaliadas pelo método PSNR. O teste aplicado foi o teste de Wilcoxon. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,9529	0,5326	<i>0,0000</i>	<i>0,0000</i>	0,6706
Sobol	0,9529	0	0,5513	<i>0,0000</i>	<i>0,0000</i>	0,6521
Uniforme	0,5326	0,5513	0	<i>0,0000</i>	<i>0,0000</i>	0,3002
Resize	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	0	<i>0,0000</i>	<i>0,0000</i>
Cropped	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	0	<i>0,0000</i>
Aleatório	0,6706	0,6521	0,3002	<i>0,0000</i>	<i>0,0000</i>	0

Fonte: Elaborada pelo autor

Aleatório e das transformações *Resize* e *Cropped*. A Tabela 11 apresenta os resultados da aplicação do teste de Wilcoxon sobre as imagens amostradas com as sequências e transformações. O teste demonstrou que os valores de qualidade obtidos da avaliação das imagens amostradas com as sequências Sobol, VDH, Uniforme e Aleatória pelo PSNR obtiveram valores próximos de qualidade. Os índices obtidos a partir das transformações *Resize* e *Cropped* são estatisticamente diferentes dos índices das demais estratégias de amostragem para as mesmas imagens.

4.4.3 Estratégia de amostragem no método *Fast-SSIM*

A amostragem de baixa discrepância foi avaliada no Fast-SSIM. O método Fast-SSIM é uma versão mais rápida do método SSIM, desta forma, a estratégia de amostragem aplicada sobre as imagens para avaliação pelo Fast-SSIM foi a mesma utilizada nas simulações do SSIM, sendo o tamanho dos blocos 32×32 pixels e a quantidade de blocos amostrados de 30 blocos de pixels para as imagens da base Live e 12 blocos de pixels para as imagens da base TID2013, que corresponde a uma taxa de amostragem aproximadamente por imagem de 7,80% e 6,25% pixels, respectivamente. A Tabela 12 apresenta os valores médios do Fast-SSIM e as correlações SROCC e LCC obtidos dos índices das imagens amostradas com as sequências VDH, Sobol, Uniforme, Aleatória e das transformações *Resize* e *Cropped*. Os valores destacados em itálico correspondem à rejeição da hipótese nula (H_0).

Tabela 12 – Valores médios de Fast-SSIM e correlações SROCC e LCC a partir das imagens amostradas usando sequências de baixa discrepância sobre a base Live. As imagens foram amostradas com 30 blocos de 32×32 pixels. O μ denota a média dos valores do Fast-SSIM, σ denota o desvio padrão e p corresponde p-value do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.

Algoritmo	Live					TID2013				
	SROCC	LCC	μ	σ	p	SROCC	LCC	μ	σ	p
Fast-SSIM (Baseline)	0,9495	0,9437	0,8358	0,1464	–	0,7039	0,7785	0,9069	0,0923	–
Fast-SSIM (VDH)	0,9441	0,9379	0,8403	0,1434	0,5282	0,6861	0,7727	0,9097	0,0906	0,4159
Fast-SSIM (Sobol)	0,9469	0,9417	0,8405	0,1427	0,5611	0,7166	0,7857	0,9096	0,0893	0,9967
Fast-SSIM (Uniforme)	0,9345	0,9279	0,8345	0,1498	0,9814	0,6145	0,7346	0,9075	0,0933	0,9777
Fast-SSIM (Resize)	<i>0,9122</i>	<i>0,6962</i>	<i>0,9599</i>	<i>0,0733</i>	<i>0,0000</i>	<i>0,6929</i>	<i>0,6791</i>	<i>0,9702</i>	<i>0,0459</i>	<i>0,0000</i>
Fast-SSIM (Cropped)	<i>0,8731</i>	<i>0,8687</i>	<i>0,7491</i>	<i>0,1807</i>	<i>0,0000</i>	<i>0,5560</i>	<i>0,6449</i>	<i>0,8437</i>	<i>0,1301</i>	<i>0,0000</i>
Fast-SSIM (Aleatório)	0,9340	0,9282	0,8425	0,1450	0,2618	0,6909	0,7684	0,9101	0,0898	0,0999

Fonte: Elaborada pelo autor

Os resultados dos experimentos demonstram que a estratégia de amostragem obteve valores de correlação SROCC próximos em relação aos valores da correlação SROCC obtidos das imagens com todos os pixels. O método de amostragem Sobol obteve o melhor desempenho da correlação SROCC na base TID2013. Os valores da correlação linear (LCC) das imagens amostradas com a sequência Sobol obtiveram melhor desempenho comparado com a correlação na avaliação utilizando as imagens completas na base Live e TID2013, os valores de LCC foram 0,8957 e 0,6912, respectivamente. Os valores do Fast-SSIM das transformações *Resize* e *Cropped* obtiveram rejeição de H_0 ($p < 0,05$), ou seja, os índices obtidos *Resize* e *Cropped* diferem dos obtidos no Fast-SSIM usando todos pixels da imagem.

O teste de Wilcoxon foi aplicado para verificar a hipótese de que os índices obtidos a partir das imagens amostradas com as sequências VDH, Sobol, Uniforme, Aleatória e transformações, *Resize* e *Cropped*, e das imagens com todos os pixels são estatisticamente diferentes. A Tabela 13 apresenta os resultados da aplicação do teste estatístico. Os resultados do teste de Wilcoxon demonstram que os valores dos índices de qualidade a partir da avaliação das imagens amostradas com as sequências VDH, Sobol, Uniforme e Aleatória não têm uma diferença significativa. Vale destacar que os valores de índices obtidos a partir da avaliação das imagens amostradas com sequência Aleatória são a média de 20 execuções. As transformações *Resize* e *Cropped* obtiveram índices significativamente diferentes das demais estratégias de amostragem.

Tabela 13 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações *Resize* e *Cropped* e avaliadas pelo método Fast-SSIM. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,9330	0,5151	<i>0,0000</i>	<i>0,0000</i>	0,6015
Sobol	0,9330	0	0,5667	<i>0,0000</i>	<i>0,0000</i>	0,5462
Uniforme	0,5151	0,5667	0	<i>0,0000</i>	<i>0,0000</i>	0,2415
Resize	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	0	<i>0,0000</i>	<i>0,0000</i>
Cropped	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	0	<i>0,0000</i>
Aleatório	0,6015	0,5462	0,2415	<i>0,0000</i>	<i>0,0000</i>	0

Fonte: Elaborada pelo autor

4.4.4 Estratégia de amostragem no método *Deep Image QuAlity Measure for Full Reference*

A estratégia de amostragem de imagens digitais foi avaliada como uma alternativa ao treinamento de modelos de redes neurais convolucionais, CNNs, para a avaliação da qualidade de imagem. O *Deep Image QuAlity Measure for Full Reference* (DIQaM-FR) (BOSSE et al., 2018) foi utilizado para avaliação da qualidade com o objetivo de investigar a estratégia de amostragem. Na fase de treinamento, O DIQaM-FR originalmente realiza a amostragem Aleatória de 32 blocos (*patches*) de tamanho 32×32 pixels para cada imagem que compõe uma época do treinamento. São selecionadas 32 imagens em cada época para o treinamento da rede, sendo 3.000 épocas no total de treinamento.

Esse comportamento do método DIQaM-FR de coletar blocos de pixels aleatoriamente durante a fase de treinamento fez surgir a hipótese de que as imagens de treinamento pudessem ser amostradas com sequências de blocos de pixels previamente definidos, evitando gerar pontos Aleatórios para a coleta de blocos em cada imagem apresentada ao método. Um ponto a observar que originalmente o método DIQaM-FR utiliza cinco etapas de agrupamento máximo (*max-pooling*) dos blocos de 32×32 pixels que daria após as cinco etapas apenas uma característica extraída na última camada. Caso o tamanho dos blocos varie, será necessário adequar a quantidade de camadas da rede. Assim, foi investigada a quantidade de camadas da rede e o tamanho dos blocos que sejam mais adequados ao

modelo para utilizar a estratégia de amostragem.

A Tabela 14 apresenta a adaptação das camadas convolucionais utilizadas no DIQaM-FR durante a etapa de treinamento, ajustando o número de camadas convolucionais conforme o tamanho do bloco de pixels para ajustar o modelo no treinamento da rede. Inicialmente, foi fixada a quantidade de 2^{14} (16.384) pixels e foram variados o tamanho e o número de blocos (*patches*) de acordo com a Tabela 14.

A análise das correlações para cada par do tamanho do bloco e do número de blocos das estratégias de amostragem VDH, Sobol, Uniforme e Aleatório foi realizada com sua respectiva quantidade de camadas convolucionais. Para cada par de configuração do tamanho do bloco e quantidade de blocos de pixels, o modelo foi treinado com 1.000 épocas no banco de dados de imagens Live. As imagens da base Live foram divididas com 60% das imagens para o treinamento, 20% para a validação e 20% para o teste. Para garantir que todos os tipos de distorções presentes na base estivesse no treinamento, esse particionamento ocorreu pela quantidade de imagens de referência, ficando com 17 imagens para treinamento, 6 imagens para validação e teste. As Figuras 20 e 21 apresentam as correlações SROCC e LCC obtidas após o treinamento do método por estratégia de amostragem e inferência das imagens do conjunto teste.

Conforme apresentado nas Figuras 20 e 21, os tamanhos de bloco 2×2 e 4×4 utilizados para amostragem das imagens de treinamento pelas sequências VDH, Sobol e Uniforme com suas respectivas camadas convolucionais correspondentes, mostraram as correlações SROCC e LCC maiores quando comparadas com outros tamanhos de bloco. Quando o tamanho do bloco é 16×16 ($M = 4$) há uma queda nas correlações. Este fato ocorre porque há uma diminuição na quantidade de blocos de 256 blocos ($M = 3$) para 64 blocos

Tabela 14 – Variação do tamanho do bloco, número de blocos e camadas convolucionais para cada tamanho de bloco na etapa de treinamento.

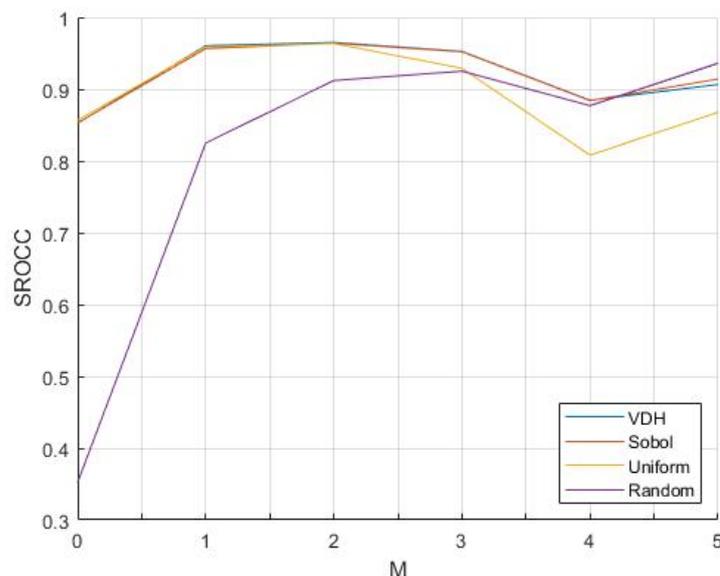
Tamanho do bloco	Qtd. de Blocos	Camadas convolucionais
2^0	128×128	conv3-32, conv3-32 (l_1)
2^1	64×64	l_1 , maxpool (l_2)
2^2	32×32	l_2 , conv3-64, conv3-64, maxpool (l_3)
2^3	16×16	l_3 , conv3-128, conv3-128, maxpool (l_4)
2^4	8×8	l_4 , conv3-256, conv3-256, maxpool (l_5)
2^5	6×6	l_5 , conv3-512, conv3-512, maxpool

Fonte: Elaborada pelo autor

($M = 4$). No caso onde $M = 5$, o modelo se aproxima da quantidade de blocos original (32 blocos) com 36 blocos amostrados por imagem com o mesmo tamanho original de 32×32 . Esse comportamento demonstra que mais blocos com tamanhos menores podem ter uma boa representatividade da imagem e, conseqüentemente, redução no número de camadas convolucionais do modelo da rede.

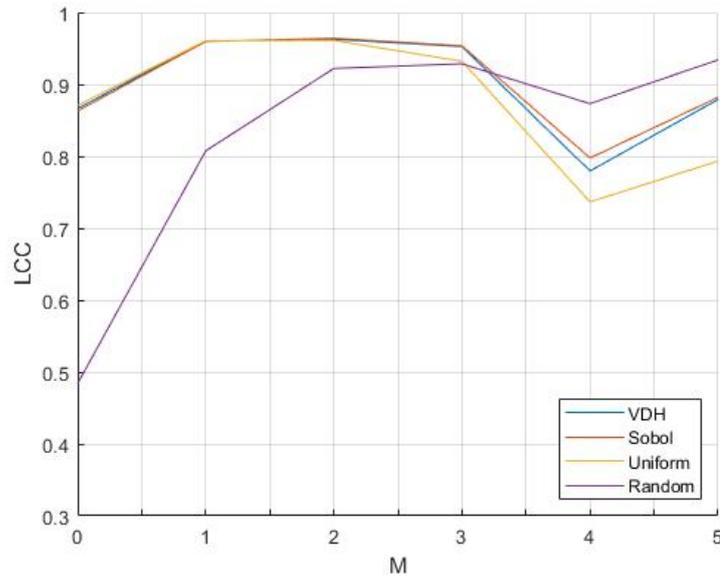
O protocolo para as simulações do método DIQaM-FR, após estabelecida a configuração de amostragem, foi comparar o desempenho do método com duas estratégias de agrupamento (*pooling*) no treinamento com as imagens completas e amostradas a partir das sequências VDH, Sobol, Uniforme e Aleatória. O tamanho do bloco de 4×4 pixels e a quantidade de 32×32 blocos foram fixados para treinamento. O modelo também foi treinado com imagens com as transformações *Resize* e *Cropped* com a mesma número de pixels ($4 \times 4 \times 32 \times 32$ pixels). O modelo DIQaM-FR foi treinado nas bases de dados Live e TID2013 com 3.000 épocas. A divisão das imagens da base Live para o teste foi de 17 imagens para treinamento, 6 imagens para validação e teste, e a TID foi 15 imagens para o treinamento e 5 para validação e teste. Os métodos de amostragem foram executados cinco vezes, e coletados os valores médios das correlações SROCC e LCC das execuções das imagens de teste das bases Live e TID2013. As imagens amostradas de pontos Aleatórios foram alteradas a cada início da fase de treinamento. Diferentemente, as imagens amostradas com as sequências VDH, Sobol e Uniforme foram amostradas apenas uma vez

Figura 20 – Análise da correlação SROCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatório variando o tamanho do bloco ($2^M \times 2^M$) nas imagens da base Live. As imagens amostradas foram avaliadas pelo método DIQaM-FR treinado com 1.000 épocas.



Fonte: Elaborada pelo autor

Figura 21 – Análise da correlação LCC dos métodos de amostragem VDH, Sobol, Uniforme e Aleatório variando o tamanho do bloco ($2^M \times 2^M$) nas imagens do banco de dados Live. As imagens amostradas foram avaliadas pelo método DIQaM-FR treinado com 1.000 épocas.



Fonte: Elaborada pelo autor

independente do início do treinamento. As sequências de baixa discrepância são geradas apenas uma vez antes do início do treinamento e são usadas em cada época.

O modelo treinado foi avaliado em dois cenários: (i) testado em imagens amostradas com tamanhos de bloco de 4×4 pixels da posição exata dos blocos das imagens de treinamento e (ii) testado em imagens de teste completas. As Tabelas 15 e 18 exibem os resultados de DIQaM-FR e WaDIQaM-FR obtidos dos bancos de dados Live e TID2013, respectivamente.

A estratégia de *pooling* médio (DIQaM-FR) treinada com as imagens amostradas com as sequências VDH e Uniforme apresentou correlação SROCC de 0,968 quando comparada ao modelo treinado com imagem completa na base Live. O método de amostragem Uniforme obteve maior valor de correlação SROCC (0,965) entre as estratégias na avaliação das imagens de teste com todos os pixels. Para a base TID2013, o DIQaM-FR treinado com imagens amostradas com a sequência VDH obteve correlação SROCC de 0,870 na avaliação das imagens amostradas com conjunto de teste. Foi observado que a estratégia ponderada de *pooling* (WaDIQaM-FR) apresentou valores de correlação SROCC e LCC próximos nas bases Live e TID2013.

As Tabelas 16 e 17 apresentam os resultados da aplicação do teste de Wilcoxon entre os pares de valores dos índices obtidos pelas estratégias de amostragem com as sequências VDH, Sobol, Uniforme, Aleatória e das transformações *Resize* e *Cropped*. Os resultados

Tabela 15 – Desempenho do modelo DIQaM-FR com duas estratégias de *pooling*. O tamanho do bloco foi de 4×4 pixels no banco de dados Live. O modelo foi treinado com imagens de amostragem e testado em imagens de amostragem e imagens completas.

Método	Live							
	Testado em blocos de 4×4 pixels				Testado na imagem completa			
	LCC	SROCC	μ_{erro}	σ_{erro}	LCC	SROCC	μ_{erro}	σ_{erro}
DIQaM-FR (Img. completa)	-	-	-	-	0,968	0,956	3,845	22,296
DIQaM-FR (Sobol)	0,962	0,967	4,208	21,892	0,960	0,962	4,593	22,017
DIQaM-FR (VDH)	0,964	0,968	3,918	21,841	0,962	0,964	4,256	21,987
DIQaM-FR (Uniforme)	0,964	0,968	3,950	21,811	0,962	0,965	4,263	21,828
DIQaM-FR (Resize)	0,960	0,948	4,732	22,330	0,867	0,871	17,739	27,913
DIQaM-FR (Cropped)	0,963	0,962	4,247	22,013	0,962	0,963	4,516	22,033
DIQaM-FR (Aleatório)	0,963	0,966	4,043	21,826	0,962	0,964	4,349	21,943
WaDIQaM-FR (Img. completa)	-	-	-	-	0,986	0,975	3,408	23,185
WaDIQaM-FR (Sobol)	0,975	0,964	3,683	22,839	0,974	0,962	3,677	22,874
WaDIQaM-FR (VDH)	0,966	0,958	4,183	22,873	0,966	0,958	4,211	22,971
WaDIQaM-FR (Uniforme)	0,971	0,960	4,160	23,202	0,975	0,965	3,991	23,345
WaDIQaM-FR (Resize)	0,952	0,937	5,849	24,167	0,886	0,820	20,193	29,710
WaDIQaM-FR (Cropped)	0,961	0,960	5,975	24,459	0,972	0,964	4,743	23,821
WaDIQaM-FR (Aleatório)	0,965	0,958	4,127	22,860	0,969	0,962	3,749	22,645

Fonte: Elaborada pelo autor

da aplicação do teste de Wilcoxon revelam que os valores de qualidade obtidos pelo método DIQaM-FR utilizando as sequências VDH e transformações *Resize* e *Cropped* não houve diferença significativa, ou seja, considerando um nível de significância de 0,05, não há diferença entre valores dos índices obtidos pelas sequências e transformações. As sequências Uniforme e Sobol não passaram no teste, demonstrando que há diferença entre os valores dos índices para o método DIQaM-FR. Para o método WaDIQaM-FR, o teste de Wilcoxon demonstrou que houve leve diferença entre os valores dos índices a partir do treinamento com as imagens amostradas com a sequência Aleatória e transformação *Resize*.

A Tabela 19 exhibe o tempo gasto no treinamento dos modelos treinados a partir das imagens com todos os pixels e com a estratégia de amostragem. A aceleração no treinamento dos métodos DIQaM-FR e WaDIQaM-FR foi de 3,61 e 3,20 vezes mais rápido comparado ao treinamento a partir das imagens com todos os pixels para a base Live. Para a base TID2013, os métodos obtiveram aceleração de 2,36 e 2,39 vezes mais rápido quando treinados a partir das imagens com todos os pixels. As imagens amostradas proporcionaram economia de tempo na fase de treinamento nos métodos de aprendizagem profunda em comparação com o treinamento com as imagens completas.

Tabela 16 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações *Resize* e *Cropped* e avaliadas pelo método DIQaM-FR. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,4956	0,0914	0,5380	0,8557	0,3990
Sobol	0,4956	0	<i>0,0138</i>	0,1642	0,6319	0,6657
Uniforme	0,0914	<i>0,0138</i>	0	0,5592	0,1065	<i>0,0254</i>
Resize	0,5380	0,1642	0,5592	0	0,4113	0,1770
Cropped	0,8550	0,6319	0,1065	0,4113	0	0,6885
Aleatório	0,3990	0,6657	<i>0,0254</i>	0,1770	0,6885	0

Fonte: Elaborada pelo autor

Tabela 17 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações *Resize* e *Cropped* e avaliadas pelo método WaDIQaM-FR. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,3240	0,2899	0,7788	0,6012	0,0659
Sobol	0,3240	0	0,8301	0,1774	0,3395	0,3015
Uniforme	0,2899	0,8301	0	0,3416	0,4541	0,2242
Resize	0,7788	0,1774	0,3416	0	0,7108	<i>0,0292</i>
Cropped	0,6012	0,3395	0,4541	0,7108	0	0,0609
Aleatório	0,0659	0,3015	0,2242	<i>0,0292</i>	0,0609	0

Fonte: Elaborada pelo autor

4.4.5 Estratégia de amostragem no método *Learned Perceptual Image Patch Similarity*

O desempenho da estratégia de amostragem foi avaliado em outro método baseado em aprendizagem profunda, o método Learned Perceptual Image Patch Similarity (LPIPS) (ZHANG et al., 2018). O LPIPS foi proposto para reproduzir o julgamento do sistema visual humano. O LPIPS foi treinado com 38.100 imagens com distorções baseadas em algorit-

Tabela 18 – Desempenho do modelo DIQaM-FR com duas estratégias de *pooling*. O tamanho do bloco foi de 4×4 pixels no banco de dados TID2013. O modelo foi treinado com imagens de amostragem e testado em imagens de amostragem e imagens completas.

Método	TID2013							
	Testado em blocos de 4×4 pixels				Testado na imagem completa			
	LCC	SROCC	μ_{erro}	σ_{erro}	LCC	SROCC	μ_{erro}	σ_{erro}
DIQaM-FR (Imagem completa)	-	-	-	-	0,807	0,770	0,624	1,054
DIQaM-FR (Sobol)	0,854	0,848	0,578	1,068	0,852	0,846	0,582	1,048
DIQaM-FR (VDH)	0,856	0,870	0,548	1,067	0,864	0,852	0,553	1,072
DIQaM-FR (Uniforme)	0,864	0,856	0,608	1,012	0,852	0,852	0,605	1,016
DIQaM-FR (Resize)	0,874	0,833	0,551	1,054	0,492	0,534	1,179	1,211
DIQaM-FR (Cropped)	0,875	0,848	0,568	1,061	0,827	0,802	0,644	1,036
DIQaM-FR (Aleatório)	0,860	0,852	0,574	1,051	0,857	0,849	0,582	1,030
WaDIQaM-FR (Imagem completa)	-	-	-	-	0,928	0,926	0,325	1,194
WaDIQaM-FR (Sobol)	0,910	0,911	0,420	1,236	0,908	0,911	0,417	1,244
WaDIQaM-FR (VDH)	0,904	0,908	0,439	1,176	0,912	0,915	0,404	1,194
WaDIQaM-FR (Uniforme)	0,903	0,905	0,462	1,220	0,913	0,917	0,445	1,216
WaDIQaM-FR (Resize)	0,892	0,888	0,472	1,288	0,627	0,647	1,179	1,211
WaDIQaM-FR (Cropped)	0,874	0,865	0,511	1,234	0,882	0,882	0,497	1,121
WaDIQaM-FR (Aleatório)	0,820	0,823	0,749	1,329	0,826	0,832	0,760	1,306

Fonte: Elaborada pelo autor

Tabela 19 – Tempo em horas de treinamento gasto para os modelos DIQaM-FR e WaDIQaM-FR.

Modelos	Live	Aceleração	TID2013	Aceleração
DIQaM-FR	26,27	-	56,30	-
DIQaM-FR (sampling)	7,27	$3,61 \times$	23,82	$2,36 \times$
WaDIQaM-FR	27,56	-	65,40	-
WaDIQaM-FR (sampling)	8,60	$3,20 \times$	27,45	$2,39 \times$

Fonte: Elaborada pelo autor

mos reais do banco de imagens *Berkeley-Adobe Perceptual Patch Similarity* (BAPPS). As imagens de referência da base BAPPS passaram por um conjunto de distorções: *autoencoding*, *denoising*, *colorization* e *super-resolution*, para formar as imagens distorcidas (ZHANG et al., 2018). O desempenho do LPIPS foi avaliado em julgamentos perceptivos de duas alternativas de escolhas forçadas, *Judgments From Two Alternatives Forced Choices* (2AFC).

O julgamento 2AFC ocorre quando uma imagem X é selecionada aleatoriamente e a ela é aplicada duas distorções para produzir as imagens X_0 e X_1 . Em seguida, as imagens que compõem a base são apresentadas a um julgador e uma pergunta é realizada “qual imagem está mais próxima da imagem original X ?”, e sua resposta (h) é registrada, $h \in \{0, 1\}$.

As imagens que compõe a base BAPPS foram obtidas aplicando distorções a partir de algoritmos reais baseados em CNN e distorções tradicionais. O grupo de imagens CNNs (4.800 imagens) foram geradas a partir de saídas de algoritmos explorando uma diversidade de tarefas, arquitetura e perdas da rede. As tarefas incluem *autoencoding*, remoção de ruído (*denoising*), colorização, e super-resolução (ZHANG et al., 2018). As imagens que compõe o grupo de imagens Tradicional (4.800 imagens) consiste em operações comuns aplicadas nas imagens de entradas. Foram utilizadas distorções fotométricas, ruído aleatório, desfoque, mudanças e corrupções espaciais e artefatos de compressão (ZHANG et al., 2018). O grupo de imagens Colorização (4.700 imagens) foram obtidas a partir de blocos de imagens da ImageNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2017) modificados pelos algoritmos pix2pix (ISOLA et al., 2016), Larson *et al.* (LARSSON; MAIRE; SHAKHNAROVICH, 2016) e Zhang *et al.* (ZHANG; ISOLA; EFROS, 2016). As imagens que compõe o grupo “Desfoque de vídeo” (9.400 imagens) foram obtidos das saídas de desfoque do *Photoshop Shake Reduction, Weighted Fourier Aggregation* (DELBRACIO; SAPIRO, 2015) e pelo método *deep video deblurring* (SU et al., 2017). As imagens modificadas por Super-resolução (10.900 imagens) foram obtidas das taxas de *upsampling* de $2\times$, $3\times$ e $4\times$ vezes com adicionais métodos *upsampling* bicúbico e quatro métodos de *deep resolution* (KIM; LEE; LEE, 2016; WANG et al., 2015; LEDIG et al., 2017; SAJJADI; SCHÖLKOPF; HIRSCH, 2017), e *Interpolação de quadros* (1.900 imagens) que foram obtidas a partir de interpolação baseada em fluxo (CE, 2009), interpolação baseada em CNN (NIKLAUS; MAI; LIU, 2017), interpolação baseada em fase (MEYER et al., 2015) e da base de imagens *Davis Middlebury* (SCHARSTEIN; SZELISKI; ZABIH, 2001), um total de imagens de 36.500 (ZHANG et al., 2018).

O protocolo das simulações do método LPIPS foi primeiramente determinar a quantidade e o tamanho dos blocos de pixels utilizado nas simulações. A escolha da quantidade de blocos e o tamanho se deu de forma empírica pela observação do desempenho da rede ao variar a quantidade de blocos entre 10 – 100 e a quantidade de pixels dos blocos em 2×2 , 4×4 , 8×8 , 16×16 , 32×32 e 64×64 pixels. A configuração que obteve melhor desempenho com a menor quantidade de pixels observado foi de 36 blocos de 8×8 pixels, as imagens de treinamento foram amostradas com essas configurações. Na fase de treinamento o modelo original utiliza apenas 5 épocas, com o menor tamanho do modelo da rede foi definido o treinamento em 10 épocas para cada base de imagens amostradas com as sequências de amostragem e transformações. Um ponto a destacar é que as imagens de validação permaneceram com o mesmo tamanho original de 256×256 pixels.

Tabela 20 – A taxa de avaliação perceptual correta do LPIPS treinado sobre as imagens distorcidas a partir da variação dos parâmetros CNN e testado nos bancos de dados de imagens de validação CNN, colorização, Tradicional, Desfocamento, Super-resolução e Interpolação de quadro. Usamos o μ para denotar a taxa média de acertos e σ para o desvio padrão.

Amostragem	CNN		Colorização		Tradicional		Desfocamento		Super-resolução		Interpolação	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
Full	64,2967	0,0493	64,5467	0,1595	69,9567	2,4988	60,6433	0,1909	71,5133	0,1662	62,7733	0,2892
VDH	64,4400	0,0265	64,4567	0,2715	72,1600	0,1179	60,9833	0,0404	71,6533	0,0551	63,0733	0,0702
Sobol	64,2333	0,0115	64,5567	0,1387	64,5267	0,1387	60,3967	0,0231	71,0067	0,0681	60,7067	0,0802
Uniforme	64,2467	0,0231	64,5567	0,1365	64,8433	0,1097	60,4467	0,0751	71,0400	0,0700	60,9833	0,1201
Resize	64,3733	0,0208	64,5500	0,3538	66,9867	0,1550	60,5133	0,0231	71,5200	0,0265	61,7700	0,0458
Cropped	64,1667	0,0404	64,8667	0,0764	66,1400	0,1253	60,6600	0,0600	71,4767	0,1415	61,5933	0,0764
Aleatório	49,9133	0,0971	50,1200	0,0656	50,2200	0,0794	49,4367	0,0971	49,6700	0,0608	49,9767	0,2797

Fonte: Elaborada pelo autor

O modelo da rede utilizado foi o AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2017). Originalmente, a LPIPS baseada na AlexNet é formada por 5 camadas convolucionais ([64, 192, 384, 256, 256]) com 3 agrupamentos (pooling) com passo de 2 (stride = 2). O AlexNet utiliza como entrada imagens de tamanho $224 \times 224 \times 3$ pixels. Entretanto, para acomodar uma imagem amostrada de tamanho menor (48×48), o modelo foi ajustado para apenas 3 camadas convolucionais ([64, 192, 384]) com 2 agrupamentos (pooling) com passo de 2 (stride = 2), eliminando as duas últimas camadas convolucionais.

A Tabela 20 exibe a porcentagem média de julgamentos corretos de LPIPS das imagens completas e amostradas com as sequências VDH, Sobol, Uniforme, e Aleatória e transformações *resize* e *cropped*. As transformações de *resize* e *cropped* usaram a mesma porcentagem do pixel da amostragem (48×48 pixels). O μ denota a taxa média correta e σ denota o desvio padrão.

O modelo treinado com as imagens amostradas com as sequências VDH, Sobol, Uniforme, e com as transformações *resize* e *cropped* obtiveram o percentual médio do julgamento correto próximo ao modelo treinado com a imagem completa na base teste CNN com as distorções *Colorization* e Desfocamento. O modelo treinado a partir das imagens amostradas com a sequência VDH obteve melhores valores quando comparado com o modelo treinado com as imagens completas nas bases de dados CNN: Tradicional, Desfocamento, *Super-Resolution* e *Frame interpolation*. Vale ressaltar que o modelo treinado com imagens amostradas com sequência VDH obteve média percentual de 72,16% superando o modelo treinado com imagens completas. Na base de teste *Colorization*, a transformação *Resize* obteve um percentual médio ligeiramente maior do julgamento correto.

Tabela 21 – Correlações SROCC e LCC do LPIPS sobre as imagens distorcidas a partir da variação dos parâmetros CNN e testado nos bancos de dados de imagens de validação CNN, colorização, tradicional, desfocamento, Super-resolução e interpolação de quadro

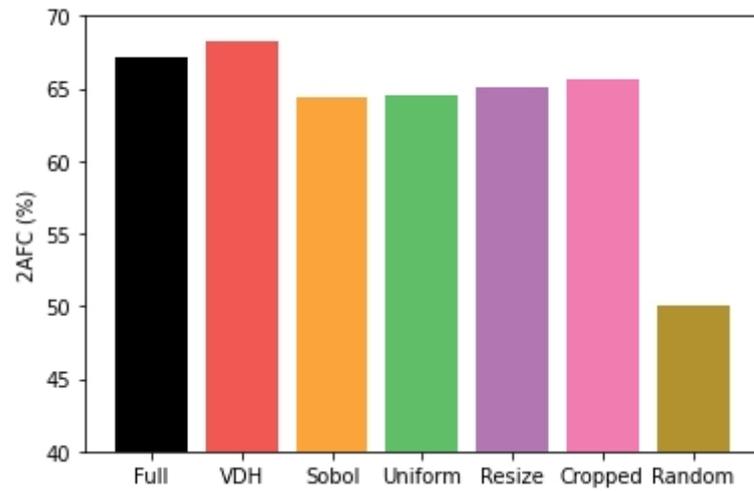
Amostragem	CNN		Colorização		Tradicional		Desfocamento		Super-resolução		Interpolação	
	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC	SROCC	LCC
Full	0,3328	0,3363	0,4846	0,4852	0,4825	0,4792	0,3641	0,3693	0,6253	0,6176	0,4128	0,4186
VDH	0,3386	0,3421	0,4801	0,4814	0,4724	0,4689	0,3699	0,3742	0,6264	0,6187	0,3973	0,4019
Sobol	0,3346	0,3383	0,4786	0,4798	0,3818	0,3784	0,3617	0,3661	0,6109	0,6033	0,3577	0,3636
Uniforme	0,3373	0,3415	0,4688	0,4697	0,3795	0,3761	0,3636	0,3680	0,6068	0,5993	0,3548	0,3608
Resize	0,3405	0,3445	0,4639	0,4644	0,4394	0,4365	0,3633	0,3679	0,6221	0,6144	0,3833	0,3884
Cropped	0,3353	0,3392	0,4882	0,4889	0,4106	0,4078	0,3671	0,3716	0,6191	0,6114	0,3728	0,3788
Aleatório	0,0014	0,0023	0,0013	0,0013	0,0055	0,055	0,0149	0,0148	0,0080	0,0088	0,0012	0,0011

Fonte: Elaborada pelo autor

A Tabela 21 apresenta as correlações SROCC e LCC do método LPIPS treinado sobre as imagens distorcidas a partir da variação dos parâmetros da CNN e testado nos bancos de dados de imagens de validação CNN, Colorização, Tradicional, Desfocamento, Super-resolução e Interpolação de quadro. A transformação *Resize* obteve maiores valores de correlações SROCC (0,3405) e LCC (0,3445) para o conjunto de imagens de validação modificadas por CNN. A transformação *Cropped* obteve melhores valores de correlações SROCC (0,4882) e LCC (0,4889) para a base colorização. A estratégia de amostragem utilizando pontos da sequência VDH obteve maiores valores de SROCC e LCC para as bases Desfocamento e Super-resolução

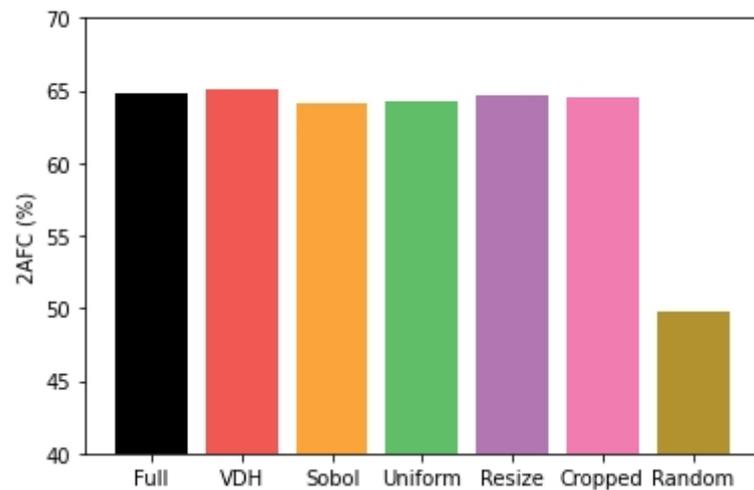
A porcentagem média de julgamento corretos foi analisada nos grupos *Algoritmos Reais* (total de imagens avaliadas foi de 26.900) e *Distorções* (total de imagens avaliadas foi de 9.600) para comparar o desempenho da estratégia de amostragem e das transformações. As Figuras 22 e 23 apresentam os resultados dos dois grupos de imagens. A LPIPS treinada com imagens amostradas a partir da sequência VDH obteve melhor porcentagem média de julgamento, 65,04% e 68,30% para os dois grupos de imagens.

Figura 22 – A taxa de sucesso dos métodos de amostragem no julgamento da escolha forçada de duas alternativas (2AFC) com LPIPS em imagens de real algoritmos (CNN e distorções tradicionais).



Fonte: Elaborada pelo autor

Figura 23 – A taxa de sucesso de métodos de amostragem em julgamento de escolha forçada de duas alternativas (2AFC) com LPIPS em imagens de distorções (super-resolução, interpolação de quadros, desfoque e colorização).

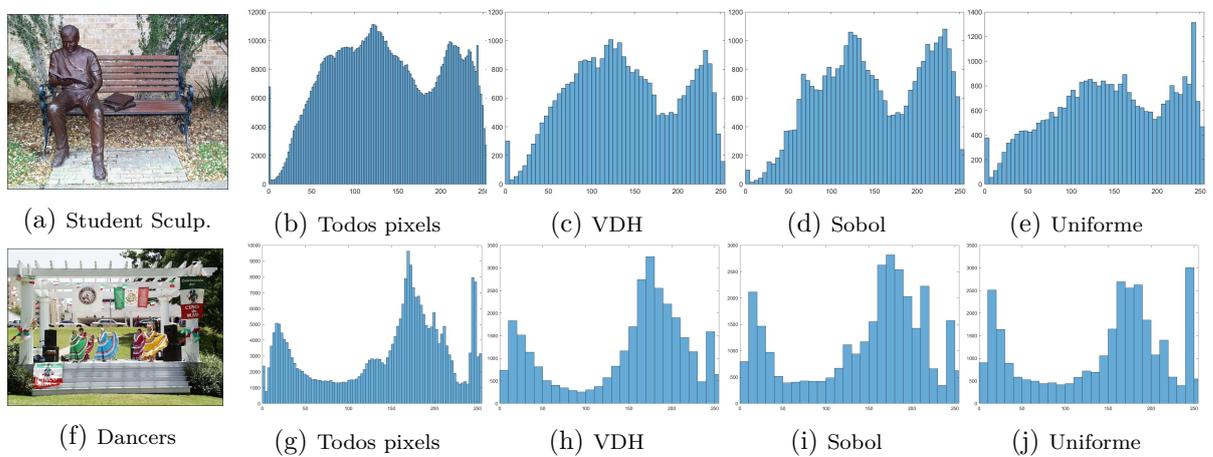


Fonte: Elaborada pelo autor

4.5 ANÁLISE DE HISTOGRAMA

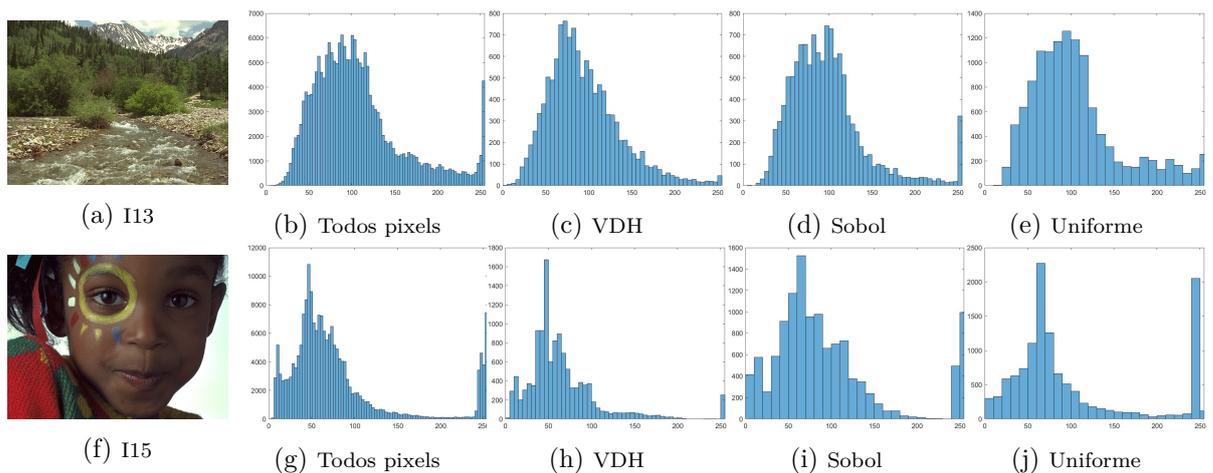
Os histogramas das imagens amostradas foram comparados com os histogramas da imagem com todos os pixels. A comparação foi realizada pela plotagem do histograma obtido após a amostragem a partir dos pontos das sequências Sobol, VDH e Uniforme. Foram selecionadas imagens da base Live e TID2013. As imagens selecionadas foram Student Sculpture e Dancers (Figuras 24a e 24f) da Live, e da base TID2013 foram I13 e

Figura 24 – Análise do histograma das imagens Student Sculpture e Dancers do banco de imagens Live. As imagens foram amostradas pelas sequências VDH, Sobol e Uniforme. A taxa de amostragem foi de 30 blocos de 32×32 pixels.



Fonte: Elaborada pelo autor

Figura 25 – Análise do histograma das imagens I13 e I15 do banco de imagens TID2013. As imagens foram amostradas pelas sequências VDH, Sobol e Uniforme. A taxa de amostragem foi de 12 blocos de 32×32 pixels.



Fonte: Elaborada pelo autor

I15 (Figuras 25a e 25f). As imagens foram escolhidas baseadas nas diferentes distribuições do histograma formados pelas imagens com todos os pixels.

A amostragem por blocos de 32×32 pixels foi utilizada para coletar pixels das imagens. A taxa de amostragem de blocos foram 30 para as imagens da base Live e 12 blocos para as imagens da base TID2013. As Figuras 24 e 25 apresentam os histogramas das imagens com todos os pixels e das imagens amostradas por cada método de amostragem.

Os histogramas obtidos a partir das estratégias de amostragem foram analisados a partir da correlação (LARSON; FARBER, 2015),

$$d(H_1, H_2) = \frac{\sum_I (H_1(I) - \overline{H_1})(H_2(I) - \overline{H_2})}{\sqrt{\sum_I (H_1(I) - \overline{H_1})^2 (H_2(I) - \overline{H_2})^2}}, \quad (4.3)$$

em que H é o histograma da imagem e $\overline{H}_k = \frac{1}{N} \sum_J H_k(J)$, e N é o total de números de pixels do histograma. A Tabela 22 apresenta o resultado da análise de correlação dos histogramas apresentados nas Figuras 24 e 25. A análise demonstra que os histogramas das imagens amostradas com as sequências VDH, Sobol e Uniforme obtêm correlação acima de 80% quando comparado com os histogramas com todos os pixels das imagens. As imagens Student Sculpture e I13 obtiveram correlação a cima de 93% a partir da amostragem com a sequência Sobol.

Tabela 22 – Análise de correlação dos histogramas obtidos a partir das imagens amostradas Student Sculpture e Dancers da base Live, I13 e I15 da base TID2012. As sequências VDH, Sobol e Uniforme foram utilizadas para amostrar as imagens. A taxa de amostragem para as imagens da base Live foi de 30 blocos e para as imagens da base TID2013 foi de 12 blocos com cada bloco de tamanho de 32×32 pixels.

Images	VDH	Sobol	Uniforme
Student Sculpture	0,9177	0,9377	0,7524
Dancers	0,8134	0,7940	0,8607
I13	0,9422	0,9688	0,8021
I15	0,8569	0,8254	0,8085

Fonte: Elaborada pelo autor

Os histogramas da imagem Student Sculpture (Figura 24a) amostrada com as sequências VDH e Sobol obtiveram semelhante distribuição dos pixels comparado o histograma formado por todos os pixels da imagem com correlação de 0,9177 e 0,9377, respectivamente.

Os histogramas obtidos a partir da amostragem da image I13 (Figura 25a) obteve correlação 0,9422 e 0,9688, respectivamente, para as estratégias de amostragem VDH e Sobol. A amostragem Uniforme obteve melhor correlação para a imagem Dancers com correlação 0,8607. Análise de histograma das imagens demonstraram que as imagens com amostragem a partir das sequências de baixa discrepância obtiveram distribuição do histograma semelhante do histograma com todos os pixels das imagens.

4.6 AMOSTRAGEM EM MÉTODOS NÃO REFERENCIADOS

A estratégia de amostragem baseada em sequência de baixa discrepância foi avaliada por métodos de IQA não referenciados. Os métodos IQA não referenciados considerados neste cenário foram Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE), Naturalness Image Quality Evaluator (NIQE) e Perception-based Image Quality Evalua-

tor (PIQUE). Os métodos NR utilizando estratégias de amostragem avaliaram as bases Live e TID213.

Cada método IQA tem uma abordagem particular para avaliar as imagens, por exemplo, tamanho da janela para filtragem. Assim, o número de blocos e o tamanho do bloco podem influenciar nos resultados dos métodos não referenciados. Portanto, os métodos NR usando a estratégia de amostragem precisam de uma configuração específica tanto no número de blocos quanto no tamanho do bloco.

A configuração utilizada em cada método não referenciado foi escolhido de forma empírica a partir da observação de diferentes configurações combinando a quantidade de blocos, variando de 10 – 100 blocos de pixels e o tamanho de blocos de pixels, variando entre 2×2 , 4×4 , 8×8 , 16×16 , 32×32 e 64×64 pixels, limitando a quantidade de pixels ao tamanho das imagens por banco de imagens. Também foi analisada a melhor estratégia de concatenação, vertical ou lado a lado.

4.6.1 Estratégia de amostragem no método *Blind/Referenceless Image Spatial Quality Evaluator*

A estratégia de amostragem utilizada foi avaliada no método não referenciado Blind/Referenceless Image Spatial Quality Evaluator (BRISQUE). A configuração que obteve melhor desempenho no BRISQUE foi com 25 blocos de 64×64 pixels. Os blocos amostrados foram concatenados na forma vertical. As bases Live e TID2013 foram utilizadas para as simulações.

A Tabela 23 exhibe as correlações SROCC e LCC do método BRISQUE avaliando as imagens amostradas nas bases Live e TID2013. Os valores destacados em itálico correspondem à rejeição da hipótese nula (H_0).

A estratégia de amostragem na base Live obteve valores de correlação SROCC e LCC próximos quando comparados às correlações obtidas avaliando as imagens completas, ou seja, utilizando todos os pixels das imagens. A amostragem Uniforme apresenta valores próximos da média e do desvio padrão para o banco de dados TID2013. As transformações *Resize*, *Cropped* e amostragem Aleatória não passaram no teste de Wilcoxon para a base TID2013 e a transformação *Resize* e amostragem Aleatória não passaram no teste Wilcoxon para a base Live.

O teste de Wilcoxon foi aplicado nos pares dos índices obtidos da avaliação das ima-

Tabela 23 – Spearman Rank-Order Correlation (SROCC) e Linear Correlation Coefficient (LCC) a partir de estratégias de amostragem aplicadas o método BRISQUE nas bases de dados Live e TID2013. Usamos o p para denotar o valor-p do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.

Algoritmo	Live					TID2013				
	SROCC	LCC	μ	σ	p	SROCC	LCC	μ	σ	p
BRISQUE (Baseline)	0,9658	0,9669	48,3796	26,0942	–	0,7136	0,7560	4,5406	0,9247	–
BRISQUE (VDH)	0,9497	0,9433	47,8423	24,0233	0,3417	0,6432	0,6625	4,5137	0,8538	0,4923
BRISQUE (Sobol)	0,9573	0,9511	46,5063	24,4603	0,1642	0,6674	0,6870	4,5495	0,8747	0,5173
BRISQUE (Uniforme)	0,9501	0,9439	48,0858	23,8013	0,3677	0,6689	0,6876	4,5387	0,8792	0,5124
BRISQUE (Resize)	<i>0,8658</i>	<i>0,8753</i>	<i>30,6863</i>	<i>20,0107</i>	<i>0,0000</i>	<i>0,4492</i>	<i>0,5355</i>	<i>4,4731</i>	<i>0,7741</i>	<i>0,0000</i>
BRISQUE (Cropped)	0,9295	0,9247	46,5468	25,2985	0,2276	<i>0,6188</i>	<i>0,6717</i>	<i>4,4592</i>	<i>0,8753</i>	<i>0,0000</i>
BRISQUE (Aleatório)	0,8635	0,8094	51,9510	23,4417	0,2801	<i>0,4782</i>	<i>0,4737</i>	<i>4,7295</i>	<i>0,7139</i>	<i>0,0000</i>

Fonte: Elaborada pelo autor

Tabela 24 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações *Resize* e *Cropped* e avaliadas pelo método BRISQUE. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,5933	0,9616	<i>0,0000</i>	0,7982	0,8929
Sobol	0,5933	0	0,5495	<i>0,0000</i>	0,7670	0,5204
Uniforme	0,9616	0,5495	0	<i>0,0000</i>	0,7796	0,9300
Resize	<i>0,0000</i>	<i>0,0000</i>	<i>0,0000</i>	0	<i>0,0000</i>	<i>0,0000</i>
Cropped	0,7982	0,7670	0,7796	<i>0,0000</i>	0	0,6852
Aleatório	0,8929	0,5204	0,9300	<i>0,0000</i>	0,6852	0

Fonte: Elaborada pelo autor

gens amostradas com as sequências VDH, Sobol, Uniforme, Aleatório e as transformações *Resize* e *Cropped* para verificar se há diferenças estatística entre os valores dos índices entre as estratégias. A Tabela 24 apresenta os resultados da aplicação do teste de Wilcoxon entre os valores de índices de qualidade obtidos a partir das estratégias de amostragem.

Os resultados do teste demonstram que não há uma diferença estatística entre os índices obtidos entre as estratégias VDH, Sobol, Uniforme, Aleatório e a transformação *Cropped*. A transformação *Resize* obteve valores de índices que diferenciam das demais estratégias. A sequência Uniforme alcançou o valor de p alto comparado com a sequência VDH indicando valores próximos dos índices entre as duas estratégias.

4.6.2 Estratégia de amostragem no método *Naturalness Image Quality Evaluator*

A amostragem utilizada nas imagens avaliadas pelo método *Naturalness Image Quality Evaluator* (NIQE) foi de 36 blocos para o Live e 25 blocos para o TID2013 com tamanho de bloco de 64×64 pixels. A escolha foi realizada de forma empírica, baseada em observação variando a quantidade e o tamanho dos blocos. Os blocos amostrados foram concatenados na forma vertical.

A Tabela 25 exhibe o resultado das correlações SROCC e LCC da estratégia de amostragem no NIQE aplicada nas bases Live e TID2013. Os valores destacados em itálico correspondem à rejeição da hipótese nula (H_0) do teste de Wilcoxon. A avaliação as imagens com amostragem de pontos Uniformes obteve melhores correlações SROCC (0,7883) e LCC (0,7673) do que a correlação obtida na avaliação das imagens com todos os pixels pelo método NIQE (Baseline) na base Live. A amostragem Aleatória e a transformação *Cropped* não passaram no teste de Wilcoxon. Os resultados das estratégias de amostragem na base TID2013 foram próximos, mas os resultados foram inferiores ao *Baseline*. As sequências Aleatórias e as transformações *Resize* e *Cropped* não passaram no teste de hipótese com $p < 0,05$.

O teste de Wilcoxon foi aplicado à análise dos pares de índices obtidos na avaliação das imagens da base Live que foram amostradas a partir das sequências VDH, Sobol, Uniforme, Aleatório e das transformações *Resize* e *Cropped*. A Tabela 26 apresenta os resultados do teste de Wilcoxon dos pares dos valores dos índices da estratégias de amostragem.

Os resultados dos testes revelam que não há uma diferença estatística entre os valores dos índices de qualidade das estratégias de amostragem. Os pares de sequências Uniforme/Sobol e *Resize*/VDH obtiveram um valor de p próximo de 1, ou seja, os valores dos índices são próximos utilizando ambas as estratégias de amostragem.

4.6.3 Estratégia de amostragem no método *Perception-based Image Quality Evaluator*

Para avaliação do método *Perception-based Image Quality Evaluator* (PIQUE), as imagens foram amostradas com 100 blocos de 16×16 pixels para as imagens da base

Tabela 25 – As correlações Spearman Rank-Order Correlation (SROCC) e Linear Correlation Coefficient (LCC) são apresentadas a partir da avaliação das imagens amostradas pelo método NIQE nas bases Live e TID2013. O p denota o valor-p do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.

Algoritmo	Live					TID2013				
	SROCC	LCC	μ	σ	p	SROCC	LCC	μ	σ	p
NIQE (Baseline)	0,7457	0,7152	3,7273	1,4170	–	0,4178	0,4722	2,9310	1,0777	–
NIQE (VDH)	0,7751	0,7566	3,8641	1,4467	0,4418	0,3785	0,4252	3,0032	1,0449	0,6232
NIQE (Sobol)	0,7770	0,7557	3,8349	1,4239	0,7978	0,3825	0,4325	2,9565	1,0223	0,5708
NIQE (Uniforme)	0,7883	0,7673	3,8658	1,4396	0,7583	0,3989	0,4472	2,9500	1,0193	0,3847
NIQE (Resize)	0,5310	0,5574	3,5568	1,0316	0,6444	0,3385	0,4248	3,2135	0,8634	0,0640
NIQE (Cropped)	0,6937	0,7137	4,2406	1,5182	0,0906	0,3513	0,3956	3,3899	1,1893	0,3447
NIQE (Aleatório)	0,7713	0,3439	11,2317	43,2544	0,5050	<i>0,3911</i>	<i>0,3927</i>	<i>3,3449</i>	<i>1,2182</i>	<i>0,0452</i>

Fonte: Elaborada pelo autor

Tabela 26 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações *Resize* e *Cropped* e avaliadas pelo método NIQE. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,5383	0,6816	0,9591	0,2382	0,8777
Sobol	0,5383	0	0,9982	0,9183	0,1119	0,6081
Uniforme	0,6816	0,9982	0	0,9992	0,1824	0,6444
Resize	0,9591	0,9183	0,9992	0	0,1510	0,5050
Cropped	0,2382	0,1119	0,1824	0,1510	0	0,6852
Aleatório	0,8777	0,6081	0,6444	0,5050	0,6852	0

Fonte: Elaborada pelo autor

Live e 32×32 pixels para as imagens da base TID2013. A escolha foi realizada de forma empírica, baseado em observação variando a quantidade e o tamanho dos blocos. Os blocos amostrados foram concatenados na forma lado a lado.

A Tabela 27 apresenta os valores de correlações SROCC e LCC das imagens avaliadas pelo PIQUE. As correlações SROCC e LCC entre os índices de avaliação pelo método PIQUE das imagens com amostragem utilizando as sequências VDH atingiu valor próximo em comparação com a correlação usando todos os pixel das imagens da base Live. A amostragem Uniforme obteve maior correlação em comparação ao valor de *Baseline* com 0,8422 e 0,8295 de correlações SROCC e LCC, respectivamente. As amostragens VDH, Sobol e Uniforme obtiveram valores de correlações próximas com uma diferença de 0,05.

Tabela 27 – As correlações Spearman Rank-Order Correlation (SROCC) e Linear Correlation Coefficient (LCC) são apresentadas a partir da avaliação das imagens amostradas pelo método PIQUE nas bases Live e TID2013 com 100 blocos de 16×16 e 32×32 pixels, respectivamente. O p denota o valor-p do teste de Wilcoxon. A linha com valores em itálico indica a rejeição do teste de hipótese.

Algoritmo	Live					TID2013				
	SROCC	LCC	μ	σ	p	SROCC	LCC	μ	σ	p
PIQUE (Baseline)	0,8398	0,8152	60,7069	20,4191	–	0,3636	0,4615	42,3253	21,2989	–
PIQUE (VDH)	0,8368	0,8259	61,6577	15,6002	0,4266	0,3131	0,4091	40,5367	17,1434	$> 0,05$
PIQUE (Sobol)	0,8294	0,8190	61,9557	15,5304	0,3560	0,3150	0,4148	40,6322	16,9863	$> 0,05$
PIQUE (Uniforme)	0,8422	0,8295	61,2824	16,2296	0,4119	<i>0,3193</i>	<i>0,3951</i>	<i>37,5519</i>	<i>18,2500</i>	$< 0,05$
PIQUE (Resize)	<i>0,7446</i>	<i>0,7390</i>	<i>52,2068</i>	<i>18,5285</i>	<i>0,0183</i>	<i>0,3582</i>	<i>0,4814</i>	<i>37,9606</i>	<i>18,9804</i>	$< 0,05$
PIQUE (Cropped)	0,8270	0,8041	61,8589	21,0693	0,6629	0,3675	0,4602	43,2066	21,7400	$> 0,05$
PIQUE (Aleatório)	<i>0,6950</i>	<i>0,6907</i>	<i>52,117</i>	<i>6,2184</i>	<i>0,0025</i>	<i>0,2415</i>	<i>0,3169</i>	<i>43,5100</i>	<i>18,2368</i>	$< 0,05$

Fonte: Elaborada pelo autor

A transformação *Resize*, e amostragem Aleatória não passam no teste Wilcoxon para o Live e nem Uniforme, *Resize* e Aleatório para o TID2013.

O teste de Wilcoxon foi aplicado aos pares de valores dos índices obtidos nas avaliações das imagens amostradas com as sequências VDH, Sobol, Uniforme, Aleatório e das transformações *Resize* e *Cropped* com o objetivo de verificar a hipótese dos valores dos índices de qualidade são estatisticamente diferentes. A Tabela 28 apresenta os resultados do teste de Wilcoxon na comparação em pares dos valores dos índices das sequências de amostragem e transformações. Os resultados demonstram que os valores de qualidade obtidos a partir das imagens amostradas com as sequências VDH, Sobol, Uniforme e a transformação *Cropped* não houve diferença significativa entre os valores de qualidade. A sequência Aleatória e a transformação *Resize* têm valores de índices diferentes das outras estratégias.

Tabela 28 – Análise estatística dos valores de índices obtidos das avaliações das imagens da base Live amostradas pelas estratégias de amostragem e transformações *Resize* e *Cropped* e avaliadas pelo método PIQUE. O teste de Wilcoxon foi aplicado para avaliar se há diferença significativa entre os índices a partir da avaliação das imagens com todos os pixels e das imagens amostradas e transformadas. O valor $p > 0,05$ do teste indica que a hipótese nula é aceita (indicando que não há diferenças entre os índices de qualidade obtidos a partir da imagem completa e os índices obtidos das imagens amostradas com as estratégias de amostragem). Os valores em itálico indicam os valores abaixo do nível de significância de 0,05.

Método de amostragem	VDH	Sobol	Uniforme	Resize	Cropped	Aleatório
VDH	0	0,9591	0,9979	0,0649	0,1741	<i>0,0103</i>
Sobol	0,9591	0	0,9989	<i>0,0455</i>	0,1824	<i>0,0103</i>
Uniforme	0,9979	0,9989	0	<i>0,0274</i>	0,2184	<i>0,0025</i>
Resize	0,0649	<i>0,0455</i>	<i>0,0274</i>	0	<i>0,0056</i>	0,8777
Cropped	0,1741	0,1824	0,2184	<i>0,0056</i>	0	<i>0,0012</i>
Aleatório	<i>0,0103</i>	<i>0,0103</i>	<i>0,0025</i>	0,8777	<i>0,0012</i>	0

Fonte: Elaborada pelo autor

4.7 ANÁLISE DE NÚMERO DE OPERAÇÕES POR PIXEL

Para avaliação do número de operações por pixel (NOPP), a estratégia de amostragem foi comparada com outro método que coleta blocos de pixels que se baseia em entropia (BRUNI; VITULANO, 2017). O trabalho de Bruni *et al.* não disponibilizou a implementação da sua estratégia online e não foi obtida resposta após contato via correio eletrônico. Desta forma, uma maneira de comparar os resultados foi realizar o comparativo do erro entre os valores disponíveis no artigo de Bruni *et al.* e os valores do SSIM das imagens amostradas e imagens completas. Isto só foi possível pela utilização das mesmas imagens com os mesmos níveis de distorções citados no artigo de Bruni *et al.*

A comparação foi realizada aplicando a amostragem VDH, Sobol e Uniforme na imagem I16 do banco de dados TID2013 (PONOMARENKO *et al.*, 2015) com as distorções: ruído de alta frequência, ruído gaussiano, borramento gaussiano, compressão JPEG, compressão JPEG2K, *Mean shift*, mudança de contraste e ruído gaussiano multiplicativo. Nas simulações, foram considerados os oito diferentes tipos de distorções e os mesmos níveis de distorção utilizados por Bruni *et al.* (BRUNI; VITULANO, 2017).

A Tabela 29 apresenta os valores do artigo de Bruni *et al.* comparados com SSIM obtidos a partir da amostragem com sequências de baixa discrepância. Os valores de SSIM de Bruni *et al.* são estimados em 30 execuções disponíveis em (BRUNI; VITULANO, 2017). Para comparar os métodos, usamos a diferença entre os valores do SSIM e o valor

de cada método ($\Delta_{x,y} = |x - y|$).

As avaliações das imagens amostradas pelo método VDH apresentam valores de SSIM próximos ao original. A amostragem VDH alcançou um erro menor que o de Bruni *et al.* em 5 dos 8 níveis de distorção de alta frequência e ruído gaussiano. O método de amostragem gasta cerca de 8 operações por pixel e requer um número menor de pixels, cerca de 6,25% da imagem completa (512×384 pixels), para realizar a estimativa do SSIM. A Tabela 30 apresenta o número de operações por pixel (nopp) usadas para cada método. O número de operações por pixel é explanado no Apêndice A.

Tabela 29 – Valores de SSIM obtidos pelo método de Bruni *et al.* e das imagens com a estratégia de amostragem utilizando as sequências Sobol, VDH e Uniforme. As simulações foram aplicados na mesma imagem I16 com seus respectivas distorções do banco de imagens TID2013. $\Delta_{x,y} = |x - y|$.

Distorções		Bruni <i>et al.</i>			Wang <i>et al.</i>	Sobol		VDH		Uniforme	
Tipo	Nível	SSIM _a	Method _b	$\Delta_{a,b}$	SSIM _c	SSIM _d	$\Delta_{c,d}$	SSIM _e	$\Delta_{c,e}$	SSIM _f	$\Delta_{c,f}$
Ruído de alta frequência	2	0,8661	0,8672	0,0011	0,8171	0,8350	0,0179	0,8166	0,0004	0,8131	0,0039
	3	0,7034	0,7086	0,0052	0,6193	0,6580	0,0387	0,6201	0,0008	0,6168	0,0025
	4	0,4829	0,4842	0,0013	0,3851	0,4408	0,0558	0,3867	0,0016	0,3879	0,0029
	5	0,2722	0,2627	0,0095	0,1943	0,2462	0,0519	0,1957	0,0014	0,2003	0,0060
Ruído gaussiano	3	0,7781	0,7821	0,0040	0,7107	0,7427	0,0320	0,7135	0,0028	0,7054	0,0053
	4	0,6614	0,6607	0,0007	0,5743	0,6230	0,0487	0,5783	0,0041	0,5674	0,0069
	5	0,5276	0,5272	0,0004	0,4313	0,4913	0,0600	0,4432	0,0118	0,4337	0,0024
	2	0,9513	0,9542	0,0029	0,8214	0,8415	0,0201	0,8238	0,0024	0,8188	0,0026
Borramento gaussiano	3	0,8805	0,8828	0,0023	0,8491	0,8176	0,0316	0,8466	0,0025	0,8747	0,0256
	4	0,7925	0,8045	0,0120	0,7473	0,7022	0,0451	0,7471	0,0002	0,7779	0,0306
	5	0,7012	0,7128	0,0116	0,6647	0,6265	0,0381	0,6741	0,0094	0,6949	0,0302
	2	0,9451	0,9448	0,0003	0,9382	0,9256	0,0126	0,9362	0,0020	0,9504	0,0123
Compressão JPEG	3	0,8891	0,8895	0,0004	0,8524	0,8495	0,0029	0,8499	0,0025	0,8883	0,0359
	4	0,7578	0,7518	0,0060	0,7249	0,7223	0,0026	0,7183	0,0066	0,7810	0,0560
	5	0,6320	0,6257	0,0063	0,6283	0,6106	0,0177	0,6475	0,0192	0,6680	0,0396
	2	0,8516	0,8553	0,0037	0,9221	0,9196	0,0024	0,9194	0,0027	0,9432	0,0212
Compressão JPEG2K	3	0,6942	0,6939	0,0003	0,6694	0,6415	0,0279	0,6742	0,0047	0,7020	0,0325
	4	0,5394	0,5529	0,0135	0,6093	0,5836	0,0257	0,6053	0,0040	0,6554	0,0460
	5	0,4799	0,4827	0,0028	0,6000	0,5672	0,0327	0,6023	0,0023	0,6359	0,0359
	2	0,9951	0,9950	0,0001	0,8119	0,7832	0,0287	0,8103	0,0016	0,8501	0,0382
Mean shift	3	0,9778	0,9703	0,0062	0,9765	0,9766	0,0001	0,9748	0,0017	0,9722	0,0043
	4	0,9620	0,9644	0,0024	0,9629	0,9647	0,0018	0,9610	0,0019	0,9651	0,0022
	5	0,8929	0,8930	0,0001	0,8866	0,8867	0,0001	0,8795	0,0071	0,8632	0,0234
	2	0,9829	0,9832	0,0003	0,9950	0,9956	0,0006	0,9951	0,0001	0,9951	0,0001
Mudança de contraste	3	0,9713	0,9711	0,0002	0,9767	0,9743	0,0025	0,9766	0,0001	0,9774	0,0007
	4	0,9349	0,9392	0,0043	0,9427	0,9401	0,0026	0,9429	0,0002	0,9455	0,0028
	5	0,8726	0,8749	0,0023	0,9021	0,8894	0,0127	0,9014	0,0007	0,9053	0,0032
Ruído gaussiano multiplicativo	3	0,8594	0,8697	0,0103	0,8069	0,8273	0,0204	0,8214	0,0145	0,8183	0,0114
	4	0,7730	0,7851	0,0121	0,7028	0,7266	0,0238	0,7228	0,0200	0,7168	0,0140
	5	0,6615	0,6627	0,0012	0,5782	0,6233	0,0451	0,6024	0,0242	0,6056	0,0274
	2	0,5376	0,5346	0,0030	0,4519	0,4959	0,0440	0,4718	0,0199	0,4747	0,0228

Fonte: Elaborada pelo autor

Tabela 30 – Número de operações por pixel (Nopp) necessárias por métrica para avaliar uma imagem de 512×384 pixels.

Métodos	Nopp	Num. de pixels	Num. de operações
SSIM	2330	512×384	1.585.109
Bruni <i>et al.</i>	10	$47 \times 17 \times 17$	135.830
Estratégia de amostragem	8,0175	$12 \times 32 \times 32$	98.520

Fonte: Elaborada pelo autor

4.8 ANÁLISE DA ACELERAÇÃO NO TEMPO NA AVALIAÇÃO DE IMAGENS AMOSTRADAS

Os tempos gastos no processo de avaliação da qualidade de imagem pelas métricas PSNR, SSIM, Fast-SSIM, BRISQUE, NIQE, PIQUE, DIQaM-FR, WaDIQaM-FR e LPIPS foram comparados na avaliação da imagem com todos os pixels (imagem completa) e da imagem utilizando a estratégia de amostragem a partir dos pontos de baixa discrepância.

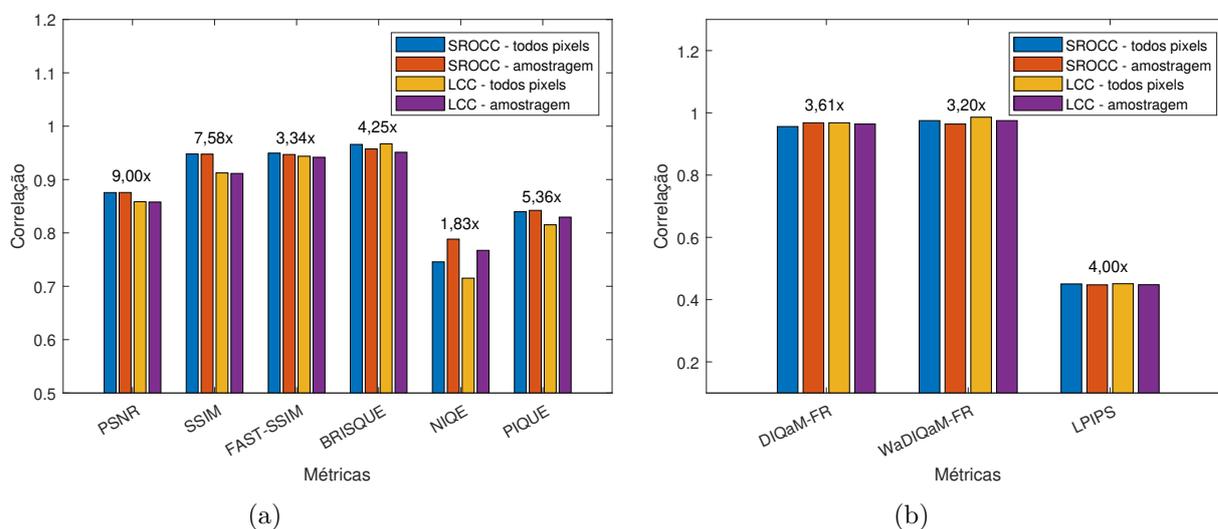
Os tempos foram calculados com a execução de vinte vezes e obtido a média dos tempos de avaliação das imagens das métricas PSNR, SSIM, Fast-SSIM, BRISQUE, NIQE e PIQUE para cada base de imagens. Os tempos das métricas baseadas em aprendizagem, DIQaM-FR, WaDIQaM-FR e LPIPS, foram comparados em relação ao processo de treinamento após cinco execuções e calculado a média dos tempos.

A Tabela 31 apresenta os tempos gastos para avaliar uma imagem completa e amostrada pelos métodos de Avaliação da Qualidade de Imagem utilizados nas simulações, bem como, a aceleração na utilização da imagem amostrada pelos métodos. Os métodos baseados em aprendizagem DIQaM-FR, WaDIQaM-FR e LPIPS obtiveram o tempo coletado em horas devido ao processo de treinamento e os demais métodos tiveram seu tempo coletado para avaliação de uma imagem em segundos.

Os tempos de avaliação demonstram que a estratégia de amostragem executa a avaliação da qualidade com menor tempo. As Figuras 26 e 27 apresentam gráficos de barras comparativo do desempenho das correlações SROCC e LCC das métricas de avaliação da qualidade de imagem utilizadas nas simulações na avaliação das imagens com todos os pixels e das imagens amostradas das bases Live e TID2013. A métrica LPIPS especificamente avaliou as imagens da base BAPPS. As métricas PSNR, SSIM e Fast-SSIM na avaliação das imagens da base Live (Figura 26a) alcançaram resultados próximos de SROCC e LCC utilizando a estratégia de amostragem na avaliação da qualidade conseguindo a aceleração na avaliação de 9 vezes, 7,58 vezes e 3,34 vezes, respectivamente, quando comparado à avaliação das imagens com todos os pixels. Para a base TID2013, estas mesmas métricas (Figura 27a) alcançaram uma aceleração no treinamento de 6 vezes, 8,28 vezes e 3,85 vezes, respectivamente.

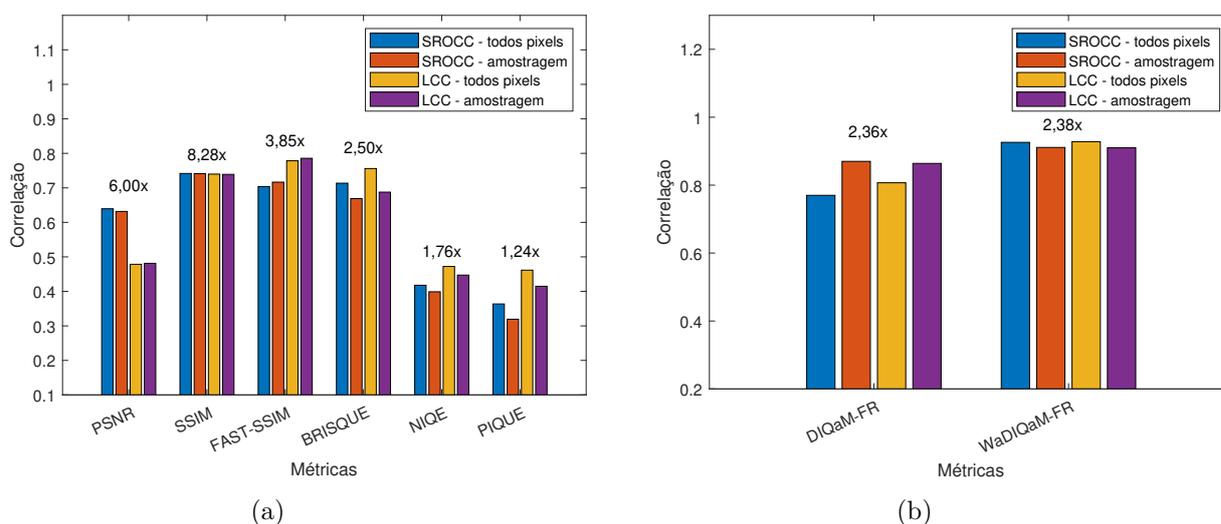
No método LPIPS, o treinamento com as imagens com amostragem conseguiu realizar o treinamento 4 vezes mais rápido comparado com as imagens completa. Vale destacar

Figura 26 – Comparativo do desempenho das correlações SROCC e LCC das métricas IQA obtidos das imagens com todos os pixels e amostradas. A métrica LPIPS avaliou as imagens da base BAPPS e as métricas PSNR, SSIM, Fast-SSIM, BRISQUE, NIQE e PIQUE avaliaram a base Live. O quantitativo no topo das correlações apresenta a aceleração obtida com a estratégia de amostragem. A Figura 26a teve seu tempo medido em segundos e a Figura 26b teve o tempo de treinamento das métricas medido em horas.



Fonte: Elaborada pelo autor

Figura 27 – Comparativo do desempenho das correlações SROCC e LCC das métricas IQA obtidos das imagens com todos os pixels e amostradas da base TID2013. O quantitativo no topo das correlações apresenta a aceleração obtida com a estratégia de amostragem. A Figura 26a teve seu tempo medido em segundos e a Figura 26b teve o tempo de treinamento das métricas medido em horas.



Fonte: Elaborada pelo autor

que para avaliação de imagem com o método PSNR, a estratégia de amostragem alcançou uma aceleração de 9 vezes em comparação a avaliação da imagem com todos os pixels.

Tabela 31 – Tempos gastos por método de IQA na avaliação das imagens com todos os pixels e com a estratégia de amostragem.

IQA	Unidade	Base	Img. completa	Img. amostrada	Aceleração
PSRN	Segundos	Live	0,0018	0,0002	9,00×
		TID2013	0,0012	0,0002	6,00×
SSIM	Segundos	Live	0,0091	0,0012	7,58×
		TID2013	0,0058	0,0007	8,28×
Fast-SSIM	Segundos	Live	0,0040	0,0012	3,34×
		TID2013	0,0027	0,0007	3,85×
BRISQUE	Segundos	Live	0,0017	0,0004	4,25×
		TID2013	0,0010	0,0004	2,50×
NIQE	Segundos	Live	0,0349	0,0190	1,83×
		TID2013	0,0219	0,0124	1,76×
PIQUE	Segundos	Live	0,0451	0,0084	5,36×
		TID2013	0,0340	0,0274	1,24×
DIQaM-FR	Horas	Live	26,27	7,27	3,61×
		TID2013	56,30	23,82	2,36×
WaDIQaM-FR	Horas	Live	27,56	8,60	3,20×
		TID2013	65,40	27,45	2,38×
LPIPS	Horas	BAPPS (CNN)	1,20	0,30	4,00×

Fonte: Elaborada pelo autor

4.9 AVALIAÇÃO DA AMOSTRAGEM EM VÍDEO

A velocidade da estratégia de amostragem foi testada em quadros de vídeo com resolução de 1920×1080 . Foi usado nas simulações 250 quadros (do quadro 451 ao 700) do conteúdo do vídeo 5 da sequência 1 do banco de dados LIVE-Netflix (BAMPIS et al., 2017; BAMPIS et al., 2016). A Figura 28 apresenta um exemplo de *frame* do vídeo utilizado nas simulações.

Figura 28 – Exemplo de um quadro do conteúdo de vídeo 5 com sequência 1 do banco de imagens LIVE-Netflix (BAMPIS et al., 2016).



Fonte: Elaborada pelo autor

Os *frames* foram amostrados a partir de pontos obtidos das sequências VDH, Sobol, Uniforme e comparados com transformações *resize* e *cropped* com o mesmo número de pixels. O número de pixels usados para amostragem foi de 35 blocos de tamanho 50×50 pixels, ou seja, cerca de 4,2% do *frame*. Os *frames* obtidos do processo de amostragem das imagens de referência e distorcida foram avaliados pelos algoritmos SSIM e Fast-SSIM.

A Tabela 32 apresenta a aceleração do SSIM com estratégia de amostragem de aproximadamente 20 vezes (de 26,74 a 526,05 fps). O SSIM com amostragem é mais rápido que o SSIM sem amostragem com um erro médio de 0,0069 usando pontos Sobol. A aceleração do Fast-SSIM com amostragem foi 6,5 vezes mais rápida (de 103,26 para 672,13 fps).

Tabela 32 – A comparação de velocidade de avaliação do SSIM e do Fast-SSIM usou uma taxa de amostragem dos blocos de 35 de 50×50 pixels sobre o vídeo 5 e na sequência 1 do banco de dados Live Netflix. O vídeo progressivo tem 250 quadros com 1920×1080 pixels de resolução.

Algoritmo	μ	σ	μ_{error}	σ_{error}	Fps
SSIM (Baseline)	0,8121	0,0289	-	-	26,74
SSIM (VDH)	0,8154	0,0277	0,0079	0,0063	526,05
SSIM (Sobol)	0,8166	0,0270	0,0067	0,0053	527,27
SSIM (Uniforme)	0,7961	0,0245	0,0185	0,0117	524,94
SSIM (Resize)	0,9661	0,0056	0,1540	0,0276	73,31
SSIM (Cropped)	0,8329	0,0436	0,0273	0,0137	148,66
SSIM (Aleatório)	0,8193	0,0371	0,0124	0,0072	516,58
Fast-SSIM (Baseline)	0,8214	0,0136	-	-	103,26
Fast-SSIM (VDH)	0,8370	0,0153	0,0155	0,0054	666,60
Fast-SSIM (Sobol)	0,8332	0,0112	0,0118	0,0055	664,96
Fast-SSIM (Uniforme)	0,8242	0,0124	0,0069	0,0056	672,13
Fast-SSIM (Resize)	0,9553	0,0070	0,1338	0,0109	162,87
Fast-SSIM (Cropped)	0,7260	0,0571	0,0955	0,0484	322,57
Fast-SSIM (Aleatório)	0,8324	0,0174	0,0124	0,0056	670,33

Fonte: Elaborada pelo autor

5 CONCLUSÃO

Este trabalho propõe uma *framework* para redução da complexidade na avaliação da qualidade de imagem (IQA) utilizando sequências de pontos predeterminados, como as sequências de baixa discrepância em métodos de referência completa (FR) e sem referência (NR). Foram analisadas duas estratégias de amostragem: pixel a pixel e por bloco de pixel. As simulações revelaram que a amostragem por bloco de pixel apresentaram melhor desempenho nos métodos de IQA investigados. A avaliação por bloco de pixels demonstrou que é possível obter uma representação da imagem com menor quantidade de pixels para avaliação da qualidade alcançando valor de índice de qualidade próximo ao índice quando avaliada a imagem com todos os pixels. A correlação entre a vizinhança dos pixels demonstrou ser importante para extração de atributos para avaliação da qualidade. A estratégia de amostragem por blocos obteve melhor resultado do que a amostragem pixel a pixel nos métodos de IQA onde a correlação da vizinhança dos pixels é importante na extração de atributos.

A análise da concatenação dos blocos de pixels lado a lado ou vertical/horizontal revelou que a forma de concatenação depende da estratégia de funcionamento de cada método de avaliação da qualidade, portanto, podendo cada estratégia de concatenação obter melhor resultado que outra.

Os métodos de IQA com ou sem a imagem de referência ao avaliar as imagens utilizando a estratégia de amostragem conseguiu obter valores de qualidade próximos a partir de uma quantidade reduzida de pixels em relação a imagem completa. O tempo de avaliação da qualidade das imagens amostradas foram menores quando comparado ao tempo gasto na avaliação da imagem com todos os seus pixels e alcançou-se uma aceleração na avaliação. Consequentemente, o número de operações foi reduzido para obter valores de índices de qualidade próximos ao valor de qualidade dos métodos de IQA avaliados com todos os pixels das imagens.

A avaliação de *frames* de vídeo por segundo (fps) mostrou que a estratégia de amostragem utilizando apenas 4,2% dos pixels obteve uma aceleração (tempo de aproximadamente 20 vezes menor) quando comparado à avaliação com todos os pixels do *frame*.

Os métodos de IQA baseados em aprendizagem treinados com imagem amostradas com as sequências de baixa discrepância alcançaram resultados de correlação semelhante

comparado com o treinamento com todos os pixel da imagem. Outro ponto a destacar é o fato de que o modelo da LPIPS baseado na AlexNet foi treinado com imagens amostradas e testado com imagens completas, ou seja, com todos os pixels que as compõem. Isto demonstra que, a partir das amostras, a rede conseguiu aprender os atributos relevantes para avaliação da qualidade quando comparado ao treinamento com a imagem completa.

5.1 DISCUSSÕES E LIMITAÇÕES

A amostragem pixel a pixel pode ser utilizada em métodos com referência completa onde a avaliação pode ser realizada pela diferença da intensidade individual de cada pixel, que é o caso do erro médio quadrático (MSE) e a relação sinal-ruído de pico (PSNR). Nos métodos de Avaliação da Qualidade de Imagem onde a correlação da vizinhança dos pixels é importante na extração de atributos, como luminância e contraste, a utilização de amostragem em bloco de pixel se torna uma estratégia mais adequada para obtenção de índices de qualidade. A quantidade de pontos para a realização da amostragem depende da estratégia de avaliação do método de IQA e do tamanho da imagem.

As sequências de baixa discrepância não apresentou uma diferença estatística relevante em comparação a sequência uniforme, mas as medidas de correção SROCC e LCC obtidas da avaliação das imagens amostradas com as sequências VDH e Sobol obtiveram melhores resultados. A avaliação da qualidade das imagens amostradas é limitada às medidas de correlação SROCC e LCC, entretanto, outras medidas de correlação são conhecidas como *Kendall Rank Correlation Coefficient* (KRCC) para serem investigadas.

O desempenho da métrica Fast-SSIM é bem superior ao das outras e com uma queda de desempenho menor. Este comportamento pode ser observado pelo motivo que o método Fast-SSIM substitui a forma de extração da luminância e contraste do SSIM para gerar uma versão mais rápida. No SSIM, a utilização de convoluções interfere nas regiões de fronteiras entre os blocos concatenados. Já no Fast-SSIM, a utilização da média e gradientes de Roberts podem ter diminuído a interferência de blocos concatenados que espacialmente poderiam estar distantes.

O treinamento dos métodos de IQA baseados em aprendizagem com imagens amostradas obtiveram avaliação das imagens equivalente quando treinado com as imagens completas, com todos os pixel. Isto permitiu diminuir a quantidade das camadas convolucionais e desta forma reduzir a quantidade de operações na fase de treinamento da

rede. Logo, o tempo de treinamento destes modelos foi reduzido alcançando o mesmo desempenho de avaliação que o obtido usando como treinamento imagens completas.

A estratégia de amostragem com pontos de sequências de baixa discrepância precisa ser investigada sobre métodos IQA de referência reduzida, além de outros tipos de modelos, como perceptuais.

A avaliação da utilização das sequência de baixa discrepância para amostragem de imagem foi limitada sobre os bancos de imagem Live e TID2013. Contudo, a investigação do desempenho da estratégia de amostragem é indispensável sobre outros bancos de imagens com diferentes tipos de distorções, como por exemplo as bases CSIQ (CHANDLER, 2010), MDID (SUN; ZHOU; LIAO, 2017), KonIQ-10k (Hosu et al., 2020) e CID2013 (VIRTANEN et al., 2015).

5.2 TRABALHOS FUTUROS

Outros cenários podem ser úteis para investigação mais aprofundada da estratégia de amostragem como trabalhos futuros:

- Avaliação em outros bancos de imagens com diferentes atributos e aplicações.
- Aplicar o *framework* proposto na avaliação da qualidade de vídeos de alta resolução (XING et al., 2022; RAO et al., 2019).
- Avaliar o desempenho dos valores de qualidade obtidos a partir das imagens amostragem com sequências de baixa discrepância em modelos de referência reduzida (*reduced Reference, RR*).
- Avaliar a estratégia de amostragem de imagens digitais utilizando pontos obtidos a partir de curvas fractais (HUANG; CHENG; HUANG, 2010) e curvas Peano (MARUSIC; KALE; TASIC, 1999).
- A reconstrução de imagem no cenário de desfocagem em rajadas de imagens necessita da seleção dos melhores quadros a partir do ordenamento destes quadros a partir de uma classificação prévia (PEÑA et al., 2020). Tanto a etapa de ordenamento como a de classificação se apresentam como cenário favorável para utilização da amostragem de imagem.

- Os métodos de classificação de imagem baseado em aprendizagem requer alta complexidade computacional na fase de treinamento, em virtude do elevado número de imagens para treinamento, por exemplo, ImageNet (RUSSAKOVSKY et al., 2015). A estratégia de amostragem é candidata como alternativa para o treinamento visando a classificação de imagens.
- A avaliação de imagens coloridas por operadores morfológicos introduz uma medida de irregularidade (índice de irregularidade) (VALLE et al., 2022). A complexidade computacional cresce na ordem de $O(m^3n^3\log(mn))$, onde m e n são as dimensões das imagens de referência e avaliada. O crescente número de operações para o cálculo do índice de irregularidade na avaliação da qualidade é um problema em que a estratégia de amostragem pode ser utilizada para redução da complexidade computacional.

5.3 ARTIGO PUBLICADO

O trabalho listado a seguir é resultado da pesquisa realizada durante o doutorado:

- “Low-discrepancy sampling for full reference image quality assessment speed up.” (GALVÃO et al., 2022).

REFERÊNCIAS

- AHUJA, N. A.; BOSE, N. K. Multidimensional generalized sampling theorem for wavelet based image superresolution. In: *2006 International Conference on Image Processing*. [S.l.: s.n.], 2006. p. 1589–1592.
- BAI, C.; REIBMAN, A. R. Image quality assessment in first-person videos. *Journal of Visual Communication and Image Representation*, v. 54, p. 123–132, 2018.
- BALANOV, A.; SCHWARTZ, A.; MOSHE, Y.; PELEG, N. Image quality assessment based on DCT subband similarity. In: *2015 IEEE International Conference on Image Processing (ICIP)*. [S.l.: IEEE, 2015. v. 2015-Decem, p. 2105–2109.
- BAMPIS, C. G.; GOODALL, T. R.; BOVIK, A. C. Sampled efficient full-reference image quality assessment models. In: *2016 50th Asilomar Conference on Signals, Systems and Computers*. [S.l.: s.n.], 2016. p. 561–565.
- BAMPIS, C. G.; LI, Z.; MOORTHY, A. K.; KATSAVOUNIDIS, I.; AARON, A.; BOVIK, A. C. *LIVE Netflix video quality of experience database*. 2016. <<http://live.ece.utexas.edu>>.
- BAMPIS, C. G.; LI, Z.; MOORTHY, A. K.; KATSAVOUNIDIS, I.; AARON, A.; BOVIK, A. C. Study of temporal effects on subjective video quality of experience. *IEEE Transactions on Image Processing*, v. 26, n. 11, p. 5217–5231, 2017.
- BECKER, S.; WIEGAND, T.; BOSSE, S. Curiously effective features for image quality prediction. In: *2021 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2021. p. 1399–1403.
- BOSSE, S.; MANIRY, D.; MULLER, K.-R.; WIEGAND, T.; SAMEK, W. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing*, v. 27, n. 1, p. 206–219, jan 2018.
- BOVIK, A. C. *The essential guide to image processing*. [S.l.]: Academic Press, 2009.
- BRATLEY, P.; FOX, B. L. Algorithm 659: Implementing Sobol’s quasirandom sequence generator. Association for Computing Machinery, New York, NY, USA, v. 14, n. 1, p. 88–100, mar. 1988.
- BRUNI, V.; VITULANO, D. An entropy based approach for ssim speed up. *Signal Processing*, v. 135, p. 198–209, 2017.
- CE, L. *Beyond pixels : exploring new representations and applications for motion analysis*. 164 p. Tese (Doutorado) — Massachusetts Institute of Technology, 06 2009.
- CERVELLERA, C.; MACCIO, D. Learning with kernel smoothing models and low-discrepancy sampling. *IEEE Transactions on Neural Networks and Learning Systems*, v. 24, n. 3, p. 504–509, 2013.
- CHAN, R.; GOLDSMITH, P. A psychovisually-based image quality evaluator for jpeg images. In: *2000 IEEE International Conference on Systems, Man and Cybernetics*. [S.l.: s.n.], 2000. v. 2, p. 1541–1546.

-
- CHANDLER, D. M. Most apparent distortion: full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, v. 19, n. 1, jan 2010.
- CHARRIER, C.; LÉZORAY, O.; LEBRUN, G. Machine learning to design full-reference image quality assessment algorithm. *Signal Processing: Image Communication*, v. 27, n. 3, p. 209–219, mar 2012.
- CHEN, B.; ZHU, L.; KONG, C.; ZHU, H.; WANG, S.; LI, Z. No-reference image quality assessment by hallucinating pristine features. *IEEE Transactions on Image Processing*, v. 31, p. 6139–6151, 2022.
- CHEN, M.-J.; BOVIK, A. C. Fast structural similarity index algorithm. In: *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. [S.l.: s.n.], 2010. p. 994–997.
- CHEN, M. J.; BOVIK, A. C. Fast structural similarity index algorithm. *Journal of Real-Time Image Processing*, v. 6, n. 4, p. 281–287, 2011.
- CONG, H.; FU, L.; ZHANG, R.; ZHANG, Y.; WANG, H.; HE, J.; GAO, J. Image quality assessment with gradient siamese network. In: *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. [S.l.: s.n.], 2022. p. 1200–1209.
- CORPUT, J. G. V. der. Verteilungsfunktionen i–ii. *Proc. Akad. Amsterdam*, v. 38, 1935.
- DALAL, I. L.; STEFAN, D.; HARWAYNE-GIDANSKY, J. Low discrepancy sequences for monte carlo simulations on reconfigurable platforms. In: *2008 International Conference on Application-Specific Systems, Architectures and Processors*. [S.l.: s.n.], 2008. p. 108–113.
- DELBRACIO, M.; SAPIRO, G. Hand-held video deblurring via efficient fourier aggregation. *IEEE Transactions on Computational Imaging*, v. 1, n. 4, p. 270–283, 2015.
- DRMOTA, M. The discrepancy of generalized van-der-Corput–Halton sequences. *Indagationes Mathematicae*, v. 26, n. 5, p. 748–759, 2015.
- EGIAZARIAN, K.; ASTOLA, J.; PONOMARENKO, N.; LUKIN, V.; BATTISTI, F.; CARLI, M. New full-reference quality metrics based on HVS. *Second International Workshop on Video Processing and Quality Metrics*, v. 4, p. 1–4, 2006.
- FAURE, H.; KRITZER, P.; PILLICHSHAMMER, F. From van der corput to modern constructions of sequences for quasi-monte carlo rules. *Indagationes Mathematicae*, v. 26, n. 5, p. 760–822, 2015. In memoriam J.G. van der Corput (1890–1975).
- FREITAS, P. G.; AKAMINE, W. Y. L.; FARIAS, M. C. Q. No-reference image quality assessment using orthogonal color planes patterns. *IEEE Transactions on Multimedia*, v. 20, n. 12, p. 3353–3360, 2018.
- FREITAS, P. G.; AKAMINE, W. Y. L.; FARIAS, M. C. Q. Referenceless image quality assessment by saliency, color-texture energy, and gradient boosting machines. *Journal of the Brazilian Computer Society*, v. 24, n. 9, 2018.

- FREITAS, P. G.; da Eira, L. P.; SANTOS, S. S.; FARIAS, M. C. Image quality assessment using bsif, clbp, lcp, and lpq operators. *Theoretical Computer Science*, v. 805, p. 37–61, 2020.
- GONZALEZ, R. C.; WOODS, R. C. *Processamento digital de imagens*. [S.l.]: Pearson, 2009.
- GU, K.; LIU, M.; ZHAI, G.; YANG, X.; ZHANG, W. Quality assessment considering viewing distance and image resolution. *IEEE Transactions on Broadcasting*, v. 61, n. 3, p. 520–531, sep 2015.
- GUO, G.; WANG, H.; YAN, Y.; ZHENG, J.; LI, B. A fast face detection method via convolutional neural network. *Neurocomputing*, v. 395, p. 128–137, 2020.
- HALTON, J. H. On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals. *Numerische Mathematik*, v. 2, p. 84–90, 1960.
- HE, X.; LIU, J.; WANG, W.; LU, H. An efficient sampling-based attention network for semantic segmentation. *IEEE Transactions on Image Processing*, v. 31, p. 2850–2863, 2022.
- HEIKH, H. R.; BOVIK, A. C.; CORMACK, L.; WANG, Z. *LIVE image quality assessment database release 2*. 2018. <<http://live.ece.utexas.edu/research/quality/subjective.htm>>.
- Hosu, V.; Lin, H.; Sziranyi, T.; Saupe, D. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, v. 29, p. 4041–4056, 2020.
- HU, B.; WANG, S.; LI, L.; LENG, J.; YANG, Y.; GAO, X. Hierarchical discrepancy learning for image restoration quality assessment. *Signal Processing*, v. 198, p. 108595, 2022.
- HUANG, C.-C.; CHENG, S.-C.; HUANG, Y. M. The design of a fractal-based number generator. In: *2010 International Workshop on Chaos-Fractal Theories and Applications*. [S.l.: s.n.], 2010. p. 435–439.
- IANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; KEUTZER, K. Squeezenet: alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size. In: *Computer Vision and Pattern Recognition Conference (CVPR)*. [S.l.: s.n.], 2017.
- ISOLA, P.; ZORAN, D.; KRISHNAN, D.; ADELSON, E. H. Learning visual groups from co-occurrences in space and time. In: *ICCV Workshop*. [S.l.: s.n.], 2016. v. 4.
- ITTI, L.; KOCH, C.; NIEBUR, E. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 20, n. 11, p. 1254–1259, 1998.
- ITU. Methodology for the subjective assessment of the quality of television pictures. *International Telecommunication Union*, BT-500-13, p. 1–48, 2012.
- KAMBLE, V.; BHURCHANDI, K. M. No-reference image quality assessment algorithms: A survey. *Optik*, v. 126, n. 11-12, p. 1090–1097, 2015.

- KIM, J.; LEE, J. K.; LEE, K. M. Accurate image super-resolution using very deep convolutional networks. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 1646–1654.
- KIM, J.; LEE, S. Deep learning of human visual sensitivity in image quality assessment framework. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.]: IEEE, 2017. p. 1969–1977.
- KIM, J.; NGUYEN, A.-D.; LEE, S. Deep cnn-based blind image quality predictor. *IEEE Transactions on Neural Networks and Learning Systems*, v. 30, n. 1, p. 11–24, 2019.
- KONG, Y.; CUI, L.; HOU, R. Full-reference iptv image quality assessment by deeply learning structural cues. *Signal Processing: Image Communication*, v. 83, p. 115779, 2020.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 60, n. 6, p. 84–90, may 2017.
- KRÖMER, P.; PLATOŠ, J.; SNÁŠEL, V. Differential evolution for the optimization of low-discrepancy generalized halton sequences. *Swarm and Evolutionary Computation*, v. 54, p. 100649, 2020.
- KUMAR, P. S.; SUBRAMANIAN, R.; SELVAM, D. T. Ensuring data storage security in cloud computing using Sobol sequence. In: *2010 First International Conference On Parallel, Distributed and Grid Computing (PDGC 2010)*. [S.l.: s.n.], 2010. p. 217–222.
- LARSON, R.; FARBER, B. *Estatística aplicada [Applied statistics]*. [S.l.]: São Paulo: Pearson Education do Brasil, 2015.
- LARSSON, G.; MAIRE, M.; SHAKHNAROVICH, G. Learning representations for automatic colorization. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. [S.l.]: Springer International Publishing, 2016. p. 577–593.
- LEDIG, C.; THEIS, L.; HUSZÁR, F.; CABALLERO, J.; CUNNINGHAM, A.; ACOSTA, A.; AITKEN, A.; TEJANI, A.; TOTZ, J.; WANG, Z.; SHI, W. Photo-realistic single image super-resolution using a generative adversarial network. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 105–114.
- LI, Y.; ZHANG, H.; CHEN, J.; SONG, P.; REN, J.; ZHANG, Q.; JIA, K. No-reference image quality assessment based on deep clustering. *Signal Processing: Image Communication*, v. 83, p. 115781, 2020.
- LIMA, E.; MADEIRO, F. Sequências de baixa discrepância aplicadas à avaliação de qualidade de imagens comprimidas. *Trends in Computational and Applied Mathematics*, v. 10, n. 2, p. 155–165, 2009.
- LING, Y.; ZHOU, F.; GUO, K.; XUE, J.-H. Assp: An adaptive sample statistics-based pooling for full-reference image quality assessment. *Neurocomputing*, v. 493, p. 568–582, 2022.

-
- LINS, R. C.; LIMA, E. A.; MELO, S. B. Estimação de movimento com uso de sequências de baixa discrepância para compressão de vídeo. *Trends in Computational and Applied Mathematics*, v. 12, n. 3, 2011.
- LIU, L.; HUA, Y.; ZHAO, Q.; HUANG, H.; BOVIK, A. C. Blind image quality assessment by relative gradient statistics and adaboosting neural network. *Signal Processing: Image Communication*, v. 40, p. 1–15, 2016.
- LIU, L.; LIU, B.; HUANG, H.; BOVIK, A. C. No-reference image quality assessment based on spatial and spectral entropies. *Signal Processing: Image Communication*, v. 29, n. 8, p. 856–863, 2014.
- LIU, M.; PO, L.-M.; XU, X.; CHEUNG, K. W.; ZHAO, Y.; LAU, K. W.; ZHOU, C. Long-range dependencies and high-order spatial pooling for deep model-based full-reference image quality assessment. *IEEE Access*, v. 8, p. 72007–72020, 2020.
- LIU, T.-j.; MEMBER, S.; LIN, W.; MEMBER, S.; KUO, C. J. Image quality assessment using multi-method fusion. *IEEE Transactions on Image Processing*, IEEE, v. 22, n. 5, p. 1793–1807, 2013.
- LV, X.; WANG, Z. J. Reduced-reference image quality assessment based on perceptual image hashing. In: *2009 16th IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2009. p. 4361–4364.
- MA, R.; HAO, Q. CS-FNet: A compressive sampling frequency neural network for simultaneous image compression and recognition. In: *2021 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI)*. [S.l.: s.n.], 2021. p. 1–6.
- MANSOURI, A.; MAHMOUDI-AZNAVEH, A. Ssvd: Structural SVD-based image quality assessment. *Signal Processing: Image Communication*, v. 74, p. 54–63, 2019.
- MARUSIC, B.; KALE, I.; TASIC, J. Image compression based on fast adaptive resampling on a hilbert-peano curve. In: *IMTC/99. Proceedings of the 16th IEEE Instrumentation and Measurement Technology Conference (Cat. No.99CH36309)*. [S.l.: s.n.], 1999. v. 1, p. 156–159 vol.1.
- MEYER, S.; WANG, O.; ZIMMER, H.; GROSSE, M.; SORKINE-HORNUNG, A. Phase-based frame interpolation for video. In: *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2015. p. 1410–1418.
- MITTAL, A.; MOORTHY, A. K.; BOVIK, A. C. Blind/referenceless image spatial quality evaluator. In: *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. [S.l.: s.n.], 2011. p. 723–727.
- 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.
- MOORTHY, A. K.; BOVIK, A. C. Blind image quality assessment: from natural scene statistics to perceptual quality. *IEEE Transactions on Image Processing*, v. 20, n. 12, p. 3350–3364, 2011.

- NAFCHI, H. Z.; SHAHKOLAEI, A.; HEDJAM, R.; CHERIET, M. Mean deviation similarity index: Efficient and reliable full-reference image quality evaluator. *IEEE Access*, v. 4, p. 5579–5590, 2016.
- NARWARIA, M.; Weisi Lin. SVD-based quality metric for image and video using machine learning. *IEEE Transactions on Systems and Cybernetics (Cybernetics)*, IEEE, v. 42, n. 2, p. 347–364, apr 2012.
- NEMA, T.; PARSAI, M. P. Reconstruction of incomplete image by radial sampling. In: *2022 International Conference on Computer Communication and Informatics (ICCCI)*. [S.l.: s.n.], 2022. p. 1–4.
- NIKLAUS, S.; MAI, L.; LIU, F. Video frame interpolation via adaptive separable convolution. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. p. 261–270.
- NIU, Y.; HUANG, D.; SHI, Y.; KE, X. Siamese-network-based learning to rank for no-reference 2d and 3d image quality assessment. *IEEE Access*, v. 7, p. 101583–101595, 2019.
- NIU, Y.; ZHONG, Y.; GUO, W.; SHI, Y.; CHEN, P. 2D and 3D image quality assessment: A survey of metrics and challenges. *IEEE Access*, IEEE, v. 7, p. 782–801, 2019.
- OPOZDA, S.; SOCHAN, A. The survey of subjective and objective methods for quality assessment of 2D and 3D images. *Theoretical and Applied Informatics*, v. 26, n. 1, p. 39–67, 2014.
- PATIL, S. B.; PATIL, S. R. Survey on approaches used for image quality assessment. *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing, ICECDS 2017*, IEEE, p. 987–991, 2018.
- PEÑA, F. A. G.; FERNÁNDEZ, P. D. M.; REN, T. I.; LEANDRO, J. de J. G.; NISHIHARA, R. M. Burst ranking for blind multi-image deblurring. *IEEE Transactions on Image Processing*, v. 29, p. 947–958, 2020.
- PO, L.-M.; LIU, M.; YUEN, W. Y. F.; LI, Y.; XU, X.; ZHOU, C.; WONG, P. H. W.; LAU, K. W.; LUK, H.-T. A novel patch variance biased convolutional neural network for no-reference image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 29, n. 4, p. 1223–1229, 2019.
- PONOMARENKO, N.; IEREMEIEV, O.; LUKIN, V.; EGIAZARIAN, K.; CARLI, M. Modified image visual quality metrics for contrast change and mean shift accounting. *11th International Conference - The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2011*, IEEE, p. 305–311, 2011.
- PONOMARENKO, N.; IEREMEIEV, O.; LUKIN, V.; EGIAZARIAN, K.; JIN, L.; ASTOLA, J.; VOZEL, B.; CHEHDI, K.; CARLI, M.; BATTISTI, F.; KUO, C.-C. J. Color image database tid2013: Peculiarities and preliminary results. In: *European Workshop on Visual Information Processing (EUVIP)*. [S.l.: s.n.], 2013. p. 106–111.

- PONOMARENKO, N.; JIN, L.; IEREMEIEV, O.; LUKIN, V.; EGI AZARIAN, K.; ASTOLA, J.; VOZEL, B.; CHEHDI, K.; CARLI, M.; BATTISTI, F.; Jay Kuo, C.-C. Image database tid2013: peculiarities, results and perspectives. *Signal Processing: Image Communication*, v. 30, p. 57–77, 2015.
- PREEDANAN, W.; KONDO, T.; BUNNUN, P.; KUMAZAWA, I. A comparative study of image quality assessment. In: *2018 International Workshop on Advanced Image Technology (IWAIT)*. [S.l.]: IEEE, 2018. p. 1–4.
- QI, Y.; JIANG, G.; YU, M.; ZHANG, Y.; HO, Y.-S. Viewport perception based blind stereoscopic omnidirectional image quality assessment. *IEEE Transactions on Circuits and Systems for Video Technology*, v. 31, n. 10, p. 3926–3941, 2021.
- RAO, R. R. R.; GÖRING, S.; ROBITZA, W.; FEITEN, B.; RAAKE, A. AVT-VQDB-UHD-1: A large scale video quality database for UHD-1. In: *2019 IEEE International Symposium on Multimedia (ISM)*. [S.l.: s.n.], 2019. p. 17–24.
- REISENHOFER, R.; BOSSE, S.; KUTYNIOK, G.; WIEGAND, T. A Haar wavelet-based perceptual similarity index for image quality assessment. *Signal Processing: Image Communication*, v. 61, p. 33–43, 2018.
- RONG, S.; WANG, Z.; WANG, J. Separated smooth sampling for fine-grained image classification. *Neurocomputing*, v. 461, p. 350–359, 2021.
- RUSSAKOVSKY, O.; DENG, J.; SU, H.; KRAUSE, J.; SATHEESH, S.; MA, S.; HUANG, Z.; KARPATHY, A.; KHOSLA, A.; BERNSTEIN, M.; BERG, A. C.; FEI-FEI, L. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision (IJCV)*, v. 115, n. 3, p. 211–252, 2015.
- SAAD, M. A.; BOVIK, A. C.; CHARRIER, C. DCT statistics model-based blind image quality assessment. In: *2011 18th IEEE International Conference on Image Processing*. [S.l.: s.n.], 2011. p. 3093–3096.
- SAAD, M. A.; BOVIK, A. C.; CHARRIER, C. Blind image quality assessment: A natural scene statistics approach in the DCT domain. *IEEE Transactions on Image Processing*, v. 21, n. 8, p. 3339–3352, 2012.
- SAJJADI, M. S. M.; SCHÖLKOPF, B.; HIRSCH, M. Enhancenet: Single image super-resolution through automated texture synthesis. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. p. 4501–4510.
- SCHARSTEIN, D.; SZELISKI, R.; ZABIH, R. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In: *Proceedings IEEE Workshop on Stereo and Multi-Baseline Vision (SMBV 2001)*. [S.l.: s.n.], 2001. p. 131–140.
- SHEIKH, H.; SABIR, M.; BOVIK, A. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, v. 15, n. 11, p. 3440–3451, 2006.
- SHEN, X.; ZENG, G.; WEI, Z. Nonuniform sampling and reconstruction for high resolution satellite images. In: *2011 International Conference on Image Analysis and Signal Processing*. [S.l.: s.n.], 2011. p. 187–191.

- SHEN, Y.; KE, J. Su-sampling based active learning for large-scale histopathology image. In: *2021 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2021. p. 116–120.
- SHI, Z.; ZHANG, J.; CAO, Q.; PANG, K.; LUO, T. Full-reference image quality assessment based on image segmentation with edge feature. *Signal Processing*, v. 145, p. 99–105, 2018.
- SOBOL, I. On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Computational Mathematics and Mathematical Physics*, v. 7, n. 4, p. 86–112, 1967.
- SU, S.; DELBRACIO, M.; WANG, J.; SAPIRO, G.; HEIDRICH, W.; WANG, O. Deep video deblurring for hand-held cameras. In: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. p. 237–246.
- SUN, W.; LIAO, Q.; XUE, J.-H.; ZHOU, F. SPSIM: A superpixel-based similarity index for full-reference image quality assessment. *IEEE Transactions on Image Processing*, IEEE, v. 27, n. 9, p. 4232–4244, sep 2018.
- SUN, W.; ZHOU, F.; LIAO, Q. Mdid: A multiply distorted image database for image quality assessment. *Pattern Recognition*, v. 61, p. 153–168, 2017.
- SÁNCHEZ-FERREIRA, C.; COELHO, L.; AYALA, H.; FARIAS, M.; LLANOS, C. Bio-inspired optimization algorithms for real underwater image restoration. *Signal Processing: Image Communication*, v. 77, p. 49–65, 2019.
- TEYTAUD, O. When does quasi-random work? In: . Berlin, Heidelberg: Springer, 2008. p. 325–336.
- THUNG, K.-H.; RAVEENDRAN, P. A survey of image quality measures. In: *2009 International Conference for Technical Postgraduates (TECHPOS)*. [S.l.]: IEEE, 2009. p. 1–4. ISBN 978-1-4244-5223-1.
- VALLE, M. E.; FRANCISCO, S.; GRANERO, M. A.; VELASCO-FORERO, S. Irregularity index for vector-valued morphological operators. *Journal of Mathematical Imaging and Vision*, Springer, p. 1–17, 2022.
- VENKATANATH, N.; PRANEETH, D.; CHANDRASEKHAR, M. B.; CHANNAPPAYYA, S. S.; MEDASANI, S. S. Blind image quality evaluation using perception based features. In: *2015 Twenty First National Conference on Communications (NCC)*. [S.l.: s.n.], 2015. p. 1–6.
- VIRTANEN, T.; NUUTINEN, M.; VAAHTERANOKSA, M.; OITTINEN, P.; HÄKKINEN, J. Cid2013: A database for evaluating no-reference image quality assessment algorithms. *IEEE Transactions on Image Processing*, v. 24, n. 1, p. 390–402, 2015.
- WANG, H.; LIL, M.; CHEN, L. No-reference image quality assessment based on multi-scale convolutional neural network assisted with visual saliency. In: *2021 International Joint Conference on Neural Networks (IJCNN)*. [S.l.: s.n.], 2021. p. 1–8.

- WANG, S.; DENG, C.; LIN, W.; HUANG, G.-b.; ZHAO, B. NMF-based image quality assessment using extreme learning Machine. *IEEE Transactions on Cybernetics*, IEEE, v. 47, n. 1, p. 232–243, jan 2017.
- WANG, S.; DENG, C.; ZHAO, B.; HUANG, G.-B.; WANG, B. Gradient-based no-reference image blur assessment using extreme learning machine. *Neurocomputing*, v. 174, p. 310–321, 2016.
- WANG, Z.; BOVIK, A. Reduced- and no-reference image quality assessment. *IEEE Signal Processing Magazine*, IEEE, v. 28, n. 6, p. 29–40, 2011.
- WANG, Z.; BOVIK, A.; SHEIKH, H.; SIMONCELLI, E. Image quality assessment: from error visibility to structural Similarity. *IEEE Transactions on Image Processing*, v. 13, n. 4, p. 600–612, apr 2004.
- WANG, Z.; BOVIK, A. C. Mean squared error: Love it or leave it? a new look at signal fidelity measures. *IEEE Signal Processing Magazine*, v. 26, n. 1, p. 98–117, 2009.
- WANG, Z.; LIU, D.; YANG, J.; HAN, W.; HUANG, T. Deep networks for image super-resolution with sparse prior. In: *2015 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2015. p. 370–378.
- WANG, Z.; SIMONCELLI, E. Translation insensitive image similarity in complex wavelet domain. In: *Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005*. [S.l.: s.n.], 2005. v. 2, p. ii/573–ii/576 Vol. 2.
- WORLEY, B. Subrandom methods for multidimensional nonuniform sampling. *Journal of Magnetic Resonance*, v. 269, p. 128–137, 2016.
- XIANG, J.; JIANG, G.; YU, M.; BAI, Y.; ZHU, Z. No-reference light field image quality assessment based on depth, structural and angular information. *Signal Processing*, v. 184, p. 108063, 2021.
- XIE, X.; CARRÉ, P.; PERRINE, C.; POUSET, Y.; ZHOU, N.; WU, J. Reduced-reference image quality metric based on statistic model in complex wavelet transform domain. *Signal Processing: Image Communication*, v. 74, p. 218–230, 2019.
- XING, F.; WANG, Y.-G.; WANG, H.; HE, J.; YUAN, J. Dvl2021: An ultra high definition video dataset for perceptual quality study. *Journal of Visual Communication and Image Representation*, v. 82, p. 103374, 2022.
- XU, T.; SHAO, Y.; WANG, Y.; QIN, H. Spatial moment pooling improves neural image assessment. In: *2022 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2022. p. 271–275.
- XUE, W.; ZHANG, L.; MOU, X.; BOVIK, A. C. Gradient magnitude similarity deviation: a highly efficient perceptual image quality index. *IEEE Transactions on Image Processing*, IEEE, v. 23, n. 2, p. 684–695, feb 2014.
- YAN, B.; BARE, B.; TAN, W. Naturalness-aware deep no-reference image quality assessment. *IEEE Transactions on Multimedia*, v. 21, n. 10, p. 2603–2615, 2019.

- YE, P.; KUMAR, J.; KANG, L.; DOERMANN, D. Unsupervised feature learning framework for no-reference image quality assessment. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2012. p. 1098–1105.
- ZHAI, G.; MIN, X. Perceptual image quality assessment: a survey. *Science China Information Sciences*, Springer, v. 63, n. 11, p. 1–52, 2020.
- ZHANG, L.; SHEN, Y.; LI, H. VSI: A visual saliency-induced index for perceptual image quality assessment. *IEEE Transactions on Image Processing*, IEEE, v. 23, n. 10, p. 4270–4281, oct 2014.
- ZHANG, L.; ZHANG, L.; BOVIK, A. C. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, IEEE, v. 24, n. 8, p. 2579–2591, 2015.
- ZHANG, L.; ZHANG, L.; MOU, X. Rfsim: A feature based image quality assessment metric using riesz transforms. In: *2010 IEEE International Conference on Image Processing*. [S.l.: s.n.], 2010. p. 321–324.
- ZHANG, R.; ISOLA, P.; EFROS, A. A. Colorful image colorization. In: LEIBE, B.; MATAS, J.; SEBE, N.; WELLING, M. (Ed.). *Computer Vision – ECCV 2016*. [S.l.]: Springer International Publishing, 2016. p. 649–666.
- ZHANG, R.; ISOLA, P.; EFROS, A. A.; SHECHTMAN, E.; WANG, O. The unreasonable effectiveness of deep features as a perceptual metric. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2018.
- ZHANG, X.; FENG, X.; WANG, W.; XUE, W. Edge strength similarity for image quality assessment. *IEEE Signal Processing Letters*, IEEE, v. 20, n. 4, p. 319–322, apr 2013.
- ZHANG, Y.; PHAN, T. D.; CHANDLER, D. M. Reduced-reference image quality assessment based on distortion families of local perceived sharpness. *Signal Processing: Image Communication*, v. 55, p. 130–145, 2017.
- ZHAO, S.; OH, S.-K.; KIM, J.-Y.; FU, Z.; PEDRYCZ, W. Motion-blurred image restoration framework based on parameter estimation and fuzzy radial basis function neural networks. *Pattern Recognition*, v. 132, p. 108983, 2022.
- ZHOU, Z.; LU, W.; YANG, J.; HE, W. No-reference image quality assessment based on neighborhood co-occurrence matrix. *Signal Processing: Image Communication*, v. 81, p. 115680, 2020.
- ZHOUWANG2011, Z.; LI, Q. Information content weighting for perceptual image quality assessment. *IEEE Transactions on Image Processing*, v. 20, n. 5, p. 1185–1198, 2011.
- ZHU, R.; ZHOU, F.; XUE, J. H. MvSSIM: A quality assessment index for hyperspectral images. *Neurocomputing*, v. 272, p. 250–257, 2018.
- ZHU, Y.; GEIß, C.; SO, E. Image super-resolution with dense-sampling residual channel-spatial attention networks for multi-temporal remote sensing image classification. *International Journal of Applied Earth Observation and Geoinformation*, v. 104, p. 102543, 2021.

ANEXO A – NÚMERO DE OPERAÇÕES POR PIXEL

O método baseado em entropia proposto por Bruni *et al.* (BRUNI; VITULANO, 2017) calcula o quantitativo de operações de multiplicação (m), de divisão (d), de soma (s) e de comparação (c) requeridas na execução do método *Structural Simirality* (SSIM). O número de operações por pixel (O) gasto na execução do SSIM é definido por

$$O_{SSIM} = O_{SSIM_p}N = (8l^2 + 18)N, \quad (\text{A.1})$$

onde O_{SSIM_p} é o número de operações requerido pelo SSIM em um pixel. Seja

$$SSIM_p = S_1S_2 = \frac{2\mu_x\mu_y + C_1}{\mu_x^2\mu_y^2 + C_1} \frac{2\sigma_{xy} + C_2}{\sigma_x^2\sigma_y^2 + C_2}, \quad (\text{A.2})$$

então

$$SSIM_p = O_{S_1} + O_{S_2} + 1m, \quad (\text{A.3})$$

quando

$$O_{S_1} = 2m + 3s + 1d + 2O_\mu \quad (\text{A.4})$$

e

$$O_{S_2} = 1m + 1d + 3s + 2O_{\sigma^2} + O_{\sigma_{xy}}, \quad (\text{A.5})$$

com $O_\mu = (l^2 - 1)s + 1d$, onde l é a janela da vizinhança de um dado pixel, $O_{\sigma^2} = (l^2 + 1)m + l^2s + 1d$, $O_{\sigma_{xy}} = (l^2 + 1)m + l^2 + 1d$. Como resultado, $O_{S_1} = (2l^2 + 1)s + 2m + 3d$, $O_{S_2} = (3l^2 + 4)m + (3l^2 + 3)s + 4d$, substituindo O_{S_1} e O_{S_2} em A.3 e somando a quantidade de operações, independente de soma, multiplicação e divisão, temos

$$O_{SSIM_p} = (8l^2 + 18). \quad (\text{A.6})$$