



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

SAULO JOSÉ DE ALBUQUERQUE SILVA

**AVALIAÇÃO AUTOMÁTICA DE QUALIDADE DE IMAGENS DE ALTA
RESOLUÇÃO SEM REFERÊNCIA**

Recife

2022

SAULO JOSÉ DE ALBUQUERQUE SILVA

**AVALIAÇÃO AUTOMÁTICA DE QUALIDADE DE IMAGENS DE ALTA
RESOLUÇÃO SEM REFERÊNCIA**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências da Computação da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de mestre em Ciências da Computação. Área de concentração: Inteligência Computacional

Orientador (a): Prof. Dr. Carlos Alexandre Barros de Mello

Recife
2022

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

S586a Silva, Saulo José de Albuquerque
Avaliação automática de qualidade de imagens de alta resolução sem
referência / Saulo José de Albuquerque Silva. – 2022.
55 f.:il., fig, tab.

Orientador: Carlos Alexandre Barros de Mello.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn,
Ciência da Computação, Recife, 2022.

Inclui referências.

1. Inteligência computacional. 2. Redes neurais. I. Mello, Carlos Alexandre
Barros de (orientador). II. Título.

006.31 CDD (23. ed.) UFPE - CCEN 2022-195

SAULO JOSÉ DE ALBUQUERQUE SILVA

**“Avaliação Automática da Qualidade de Imagens de Alta
Resolução Sem Referência”**

Dissertação de Mestrado apresentada
ao Programa de Pós-Graduação em
Ciência da Computação da Universidade
Federal de Pernambuco, como requisito
parcial para a obtenção do título de
Mestre em Ciência da Computação.

Aprovado em: 06/09/2022.

BANCA EXAMINADORA

Prof. Dr. Cleber Zanchettin
Centro de Informática /UFPE

Prof. Dr. Rafael Galvão de Mesquita
IFPE / Campus Garanhuns

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

LISTA DE FIGURAS

Figura 1 – Foto escura na floresta, que passa a sensação de suspense e mistério. Por padrões usuais, essa foto tem baixa qualidade, mas a intenção era de gerar uma foto escura mesmo.	17
Figura 2 – Visão geral do OCPP no plano de cores HSV	22
Figura 3 – Planos XY, XZ e YZ respectivamente	22
Figura 4 – Fluxo fim-a-fim do DIQA	28
Figura 5 – Arquitetura DIQA para previsão de qualidade de imagem	29
Figura 6 – Arquitetura do modelo EDSR	34
Figura 7 – Blocos residuais da ResNet (esquerda) e EDSR (direta)	35
Figura 8 – Imagens tiradas em ambiente externo com dimensões de 4640x3472 pixels.	46
Figura 9 – Fotografias tiradas com variação de zoom com dimensões de 4640x2610 pixels. Nas imagens com zoom, é percebido uma redução na qualidade. Logo, (b) tem menor qualidade visual que (a) e (c) tem menor qualidade visual que (b)	46
Figura 10 – Fotografias tiradas em ambiente interno e com zoom com dimensões de 4640x3472 pixels. Em (a) o objeto em foco está mais distante logo, existe mais área borrada que em (b).	47
Figura 11 – Imagens avaliadas com o escores incoerentes pelo DIQA, todas as imagens têm dimensões de 4624x2136 pixels. Super exposição e borramento do céu em (a).	49
Figura 12 – Imagens avaliadas com o escore 0,9972 pelo OCPP, todas as imagens têm dimensões de 4640x3472 pixels. Alto nível de borramento na parte superior de (a). Borramento em todo o reflexo no lado esquerdo de (b).	49

LISTA DE TABELAS

Tabela 1 – Resumo das características extraídas pelo BRISQUE	26
Tabela 2 – Configurações de hardware utilizadas na execução dos experimentos . .	39
Tabela 3 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência quando treinado e aplicado nas bases LIVE e KonIQ-10K.	40
Tabela 4 – Escores preditos pelo algoritmo DIQA em comparação os escores definidos por avaliadores humanos: exemplos da base LIVE IQA.	40
Tabela 5 – Escores preditos pelo algoritmo DIQA em comparação os escores definidos por avaliadores humanos: exemplos da base KonIQ-10K.	41
Tabela 6 – Resultados da avaliação quando o DIQA é treinado em uma base de imagens e aplicado em outra.	41
Tabela 7 – Experimento com variação de borramento: foi aplicado um filtro Gaussiano de borramento, na imagem original, com diferentes níveis. A imagem da linha inferior está mais borrada que a da linha superior da tabela.	42
Tabela 8 – Experimento com variação perda no armazenamento em JPEG: a imagem da linha superior foi salva com 80% de perda e a da linha inferior com 90% sobre 80% de perda.	43
Tabela 9 – Experimento com variação de nível de brilho.	43
Tabela 10 – Experimento com alta saturação do canal vermelho.	43
Tabela 11 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência utilizando a base LIVE IQA.	44
Tabela 12 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência utilizando a base LIVE in the Wild.	44
Tabela 13 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência utilizando a base KonIQ-10K.	45
Tabela 14 – Tempo médio para avaliar uma imagem de alta resolução	45
Tabela 15 – Escores das imagens da Figura 8	46
Tabela 16 – Escores das imagens da Figura 9	47
Tabela 17 – Escores das imagens da Figura 10	48
Tabela 18 – Escores das imagens da Figura 9 atribuídos pelo IL-NIQE e NBIQA. .	50

Dedico esta dissertação a toda minha família, aos amigos e professores que me deram o apoio necessário para chegar até aqui.

AGRADECIMENTOS

No mestrado e no curso de uma pesquisa, sabemos que podem existir momentos de dúvidas, incertezas e um pouco de angústia. Quando entrei, ninguém poderia imaginar que passaríamos por uma pandemia e que esses sentimentos seriam intensificados. Mas, durante a caminhada, tive suporte e apoio de pessoas incríveis que me auxiliaram para que a caminhada fosse possível.

Primeiramente, agradeço a Deus e minha família, especialmente aos meus pais, esposa e irmão, por me incentivarem e sempre acreditarem no meu potencial.

Aos meus amigos, sempre dispostos a ouvir minhas incertezas e questionamentos e ajudar no que fosse preciso.

Ao professor Carlos, meu orientador, por todo conhecimento, paciência e compreensão nas reviravoltas que a vida deu no percurso do mestrado.

À Universidade Federal de Pernambuco e ao Centro de Informática, pelo acolhimento e possibilidade de sempre aprofundar meus conhecimentos.

Ao pessoal da Motorola, principalmente, a Paulo Guarda, pelo apoio, ideias, sugestões e desafios lançados.

RESUMO

Um dos grandes desafios nas áreas de visão computacional e processamento de imagens é desenvolver métodos para modelar corretamente o sistema visual humano (SVH). Um dos processos executados pelo sistema visual humano é a avaliação da qualidade de uma imagem, processo complexo por envolver elementos objetivos e subjetivos. Dentre as abordagens de avaliação automática da qualidade de imagens da literatura, neste trabalho é utilizada a avaliação da qualidade de imagens sem referência (*No-reference Image Quality Assessment* - IQA), um tipo de avaliação que utiliza como dado de entrada apenas a imagem que será avaliada. Os modelos IQA podem ter um propósito específico ou ser de uso geral. Esses modelos possuem diversas aplicações, que vão desde calibrar parâmetros de câmeras até medir a degradação de um canal de comunicação por onde passam imagens. Assim, é de grande importância desenvolver modelos cada vez mais precisos e eficientes. Nos últimos anos, houve um avanço na tecnologia das câmeras de dispositivos móveis, possibilitando a captura de imagens de alta resolução. Nesse cenário, utilizamos as imagens geradas por tais dispositivos para avaliar o comportamento de métodos de avaliação automática da qualidade de imagens propostos na literatura quando submetidos a imagens de alta resolução. Dentre os métodos existentes destacamos: (1) BRISQUE, (2) OCPP, e (3) DIQA. O modelo DIQA é baseado em inteligência artificial e, com isso, foram utilizadas para treinamento as bases de dados públicas e amplamente conhecidas: LIVE IQA, LIVE *in the Wild* e KonIQ-10K. Essas bases públicas possuem imagens de qualidade variada e baixa resolução. Até onde sabemos, não existem bases públicas com imagens de alta resolução rotuladas com uma nota de qualidade. O objetivo deste trabalho é avaliar os modelos em referência existentes na literatura, utilizando imagens de alta resolução. Como utilizamos o SVH como inspiração, serão considerados na avaliação das imagens critérios objetivos e subjetivos. Visando melhorar os resultados desse cenário, foram feitas modificações na arquitetura convolucional e no pré-processamento das imagens do DIQA. Todos os modelos foram avaliados utilizando imagens de alta resolução de cunho autoral. Os resultados da aplicação desses modelos a tais imagens foram bastante incoerentes, todos os modelos apresentaram algum tipo de inconsistência como é apresentado nesta Dissertação. Sugere-se que pesquisas posteriores abordem modelos de aprendizagem profundo que visem corrigir os problemas encontrados no DIQA para torná-lo mais eficiente em termos de acurácia e processamento.

Palavras-chave: avaliação da qualidade de imagem; imagens de alta resolução; redes neurais convolucionais.

ABSTRACT

A great challenge in computer vision and image processing is to develop methods to model correctly the human visual system (HVS). One of the processes that runs in the human visual system is the evaluation of the quality of an image, a complex task because it involves objective and subjective elements. Among the automatic evaluation approaches of image quality in the literature, this work uses the evaluation of image quality without reference (No-reference Image Quality Assessment - IQA), a type of evaluation that uses as input only the image to be evaluated. IQA models can be of specific or general purpose. These models have several applications, ranging from calibrating camera parameters to measuring the degradation of a communication channel through which images can pass. Thus, it is of great importance to develop accurate and efficient models. In recent years, there have been advances in the technology of mobile device cameras, enabling the capture of high resolution images. In this scenario, we used the images generated by such mobile devices to evaluate the behavior of automatic IQA methods proposed in the literature when applied to high resolution images. Several models were found in the literature as: (1) BRISQUE, (2) OCPP, and (3) DIQA. For DIQA, which is based on artificial intelligence, public databases well-known were used for training, such as: LIVE IQA, LIVE in the Wild and KonIQ-10K. These public databases have images of varying quality and low resolution. As far as we know, there are no public databases with high resolution images labeled with a quality grade. The objective of this work is to evaluate the No-reference IQA models existing in the literature using high-resolution images. As we use the HVS as inspiration, objective and subjective criteria will be considered in the evaluation of the images. In order to improve the results of this scenario, modifications were made in the convolutional architecture and in the image pre-processing step of DIQA. All models were evaluated using a high-resolution images dataset created in this work. The results of applying these models to such images were quite inconsistent; all models showed some kind of inconsistency as shown in this Dissertation. It is suggested that further research addresses deep learning models that aim to correct the problems found in DIQA to make it more efficient in terms of accuracy and processing time.

Keywords: image quality assessment; high-resolution images; convolutional neural networks.

LISTA DE ABREVIATURAS E SIGLAS

ADAM	Otimizador de Estimação de Momento Adaptativo
AGGD	Distribuição Gaussiana Generalizada Assimétrica
BN	<i>Batch Normalization</i>
CNN	Redes Neurais Convolucionais
DIQA	<i>Deep Image Quality Assessment</i>
DS	<i>Distortion Specific</i>
EDSR	<i>Enhanced Deep Super Resolution Network</i>
FR	<i>Full Reference</i>
GAP	<i>Global Average Pooling</i>
GB	Borramento Gaussiano
GGD	Distribuição Gaussiana Generalizada
GPIQA	<i>General Purpose IQA</i>
GPU	Unidade de processamento gráfico
IQA	<i>Image Quality Assessment</i>
KROCC	<i>Kendall Rank Order Correlation Coefficient</i>
LBP	<i>Local Binary Pattern</i>
LIVE	Laboratório de Engenharia de Imagem e Video
MOS	<i>Mean Opinion Score</i>
MSCN	Média Subtraída do Contraste Normalizado
NBIQA	<i>Novel Blind Image Quality Assessment</i>
NIQE	<i>Natural Image Quality Evaluator</i>
NNR	<i>Nearest Neighbor Regression</i>
NR	<i>No-Reference</i>
NRIQA	<i>No-Reference Image Quality Assessment</i>
NSS	<i>Natural Scene Statistics</i>
OCP	<i>Orthogonal Color Planes Pattern</i>
PLCC	<i>Pearson Linear Correlation Coefficient</i>
RELU	Unidade Linear Retificada
RF	<i>Random Forest</i>
RMSE	<i>Root Mean Square Error</i>

RR	<i>Reduce Reference</i>
SROCC	<i>Spearman Rank Order Correlacionados Coefficient</i>
SVH	Sistema Visual Humano
SVR	<i>Support Vector Regression</i>

LISTA DE SÍMBOLOS

Σ	Somatório
\int	Integral
ζ	Letra grega minúscula zeta
\in	Pertence
\geq	Maior ou Igual
δ	Delta
θ	Teta
σ	Sigma
μ	Mi
∞	Infinito
ρ	Rô
π	Pi
τ	Tau
ω	Ômega
α	Alfa
β	Beta
Γ	Gama
ϵ	Épsilon
η	Eta
θ	Teta

SUMÁRIO

1	INTRODUÇÃO	15
1.1	OBJETIVOS	17
1.2	ORGANIZAÇÃO DO TRABALHO	18
2	ESTADO DA ARTE PARA AVALIAÇÃO SEM REFERÊNCIA DA QUALIDADE DE IMAGENS	19
2.1	ALGORITMOS ESTATÍSTICOS	19
2.1.1	Orthogonal Color Planes Patterns - OCPP	20
2.1.1.1	Descritor OCPP	20
2.1.1.2	Modelo IQA	22
2.1.2	Blind/Referenceless image spatial quality evaluator (BRISQUE)	22
2.2	ALGORITMO BASEADO EM APRENDIZAGEM PROFUNDA	26
2.2.1	Deep Image Quality Assessment - DIQA	27
2.2.1.1	Arquitetura do Modelo	27
2.2.1.2	Normalização da Imagem	28
2.2.1.3	Avaliação subjetiva	29
2.2.1.4	Treinamento	29
3	VARIAÇÕES PROPOSTAS NO MODELO DIQA	31
3.1	PROBLEMA	31
3.2	PROPOSTAS DE APRIMORAMENTO DO DIQA	31
3.2.1	Normalização (pré-processamento)	31
3.2.2	Super-resolução	32
4	EXPERIMENTOS	35
4.1	MEDIDAS PARA AVALIAÇÃO	35
4.1.1	Pearson Linear Correlation Coefficient - PLCC	35
4.1.2	Spearman Rank Order Correlation Coefficient - SROCC	35
4.1.3	Kendall's Rank Order Correlation Coefficient - KROCC	36
4.1.4	Root Mean Square Error - RMSE	36
4.2	BASES	36
4.2.1	Base LIVE IQA	37
4.2.2	Base LIVE In the Wild	37
4.2.3	KonIQ-10k	37
4.3	AMBIENTE	38

4.4	EXPERIMENTOS COM O DIQA	38
4.5	RESULTADOS	42
4.6	OUTROS EXPERIMENTOS	48
5	CONCLUSÃO	50
5.1	TRABALHOS FUTUROS	50
	REFERÊNCIAS	52

1 INTRODUÇÃO

Um dos maiores desafios nos campos de visão computacional e processamento de imagem é desenvolver métodos para modelar corretamente o sistema visual humano. Por sistema visual, entendemos todos os processos relacionados à visão que vão desde a captura da imagem até o reconhecimento completo dela. Dentre esses processos, está a avaliação da qualidade de uma imagem.

A avaliação da Qualidade da Imagem busca sistematizar medidas que definam automaticamente a qualidade da imagem percebida, sendo a qualidade da imagem uma propriedade que mensura o nível de degradação percebida em uma imagem.

No caso de imagens digitais, a qualidade pode ser afetada e a imagem sofrer distorções em razão de diversos fatores durante os processos aos quais elas são submetidas, como transmissão, processamento, compressão e armazenamento. Além desses, algumas distorções podem surgir do próprio processo de fotografar como, por exemplo, um movimento involuntário ao tirar uma fotografia que pode gerar um borramento de deslocamento na imagem.

É importante salientar que a subjetividade é um aspecto importante e uma dificuldade enfrentada quando se trata da avaliação da qualidade de uma imagem, pois uma foto pode estar apropriada e servir bem ao objetivo de quem a vê. Por exemplo, a Figura 1 apresenta uma foto escura que intenciona passar uma sensação de mistério. Contudo, tal imagem ao ser avaliada no computador será mensurada como muito escura, já que a avaliação não terá como modelar a experiência humana ou a intenção existente ao criar a imagem. sjas

A avaliação da qualidade da imagem pode ocorrer seguindo uma dentre duas possíveis classificações: (1) classificação quanto ao critério de avaliação e (2) classificação de acordo com a presença ou não alguma referência comparativa. Quando se fala de classificação quanto ao critério de avaliação há duas possibilidades: a primeira é a subjetiva, que está relacionada aos critérios pessoais dos sujeitos que não são mensuráveis; a segunda é a objetiva que pode ser mensurada quantitativamente e é possível de ser reproduzida.

No que diz respeito à classificação de acordo com a presença ou não de uma imagem de referência há três tipos. A classificação com referência completa *Full Reference* (FR) que acontece quando se compara duas imagens. Métodos FR, em geral, são rápidos e eficientes, sendo úteis, por exemplo, na avaliação do resultado de transmissões de imagens devido ao uso de codificadores. A grande vantagem desses métodos é que podemos comparar a imagem original com a imagem distorcida. Contudo, a imagem original (sem distorção) nem sempre está disponível.

Existem também os métodos sem referência *No-Reference* (NR) ou *Blind* em que a avaliação acontece apenas em uma única imagem. Esses são bastante úteis quando não temos informações da imagem original (sem distorção). Em geral, esses métodos são

Figura 1 – Foto escura na floresta, que passa a sensação de suspense e mistério. Por padrões usuais, essa foto tem baixa qualidade, mas a intenção era de gerar uma foto escura mesmo.



Fonte: QUOTEV (2022)

custosos.

Por fim, existem os métodos de referência reduzida *Reduce Reference* (RR) que utilizam partes da imagem original para avaliar a imagem alvo. Tais métodos são considerados um meio termo, já que utilizam apenas parte da imagem original; em geral, buscam por degradações específicas.

Existem aplicações práticas para todos esses modos. No caso de telefones celulares e outros dispositivos móveis, os métodos FR podem ser usados para comparação entre as imagens capturadas por telefones diferentes com configurações semelhantes. Os métodos NR podem avaliar a qualidade da imagem a ser capturada e fazer ajustes automáticos nos parâmetros da câmera (correção de iluminação, por exemplo). Já os métodos RR podem

ser usados mais diretamente na análise de vídeo, onde um quadro anterior pode ser usado para analisar um quadro atual.

Em um ambiente de teste, os métodos FR e NR podem ser facilmente aplicados em diversas aplicações na análise de imagens fotografadas pelos telefones celulares, como por exemplo a criação de testes automáticos com análises mais objetivas. Este ponto é interessante porque a análise atual às vezes é feita por um operador humano. Mesmo no mesmo cenário (considerando a luminosidade, o tipo de tela e assim por diante), cada pessoa pode ter sua própria definição pessoal de qualidade. Essa diferença na percepção das imagens é um dos motivos para o desenvolvimento de métodos objetivos de avaliação da qualidade. Em uma aplicação, onde várias imagens diferentes devem ser analisadas em muitas condições diferentes, é difícil ter resultados que não sejam afetados por aspectos humanos como fadiga, por exemplo. Ainda mais, é muito trabalhoso replicar experimentos qualitativos e aplicá-los em um grande conjunto de dados.

Existem diversas medidas que podem ser utilizadas durante a avaliação de qualidade de imagem, uma delas é *Mean Opinion Score* (MOS) (ou Escore de Opinião Médio) que é considerada uma avaliação subjetiva em que pessoas observam uma imagem e atribuem valores, indicando se a imagem tem boa qualidade ou não. Por exemplo, um MOS de um (1) a cinco (5), pode ter (1) para avaliar uma imagem ruim e (5) para avaliar uma imagem como excelente. Em alguns casos, essa escala é convertida para o intervalo [0-100] para facilitar a interpretação do resultado, sendo 0 para considerar uma imagem ruim e 100 uma imagem excelente. Existem outras formas de mudança na escala MOS, mas não são abordadas nesta dissertação, por fugirem do nosso escopo. Essa mensuração, contudo, pode demandar bastante tempo por ser muito trabalhosa, ainda mais quando se trata de bases de imagens muito grandes.

Diante do exposto, este estudo busca explorar e avaliar a capacidade e a precisão de algoritmos sem referência para avaliação automática de qualidade de imagens quando aplicados a imagens de alta resolução. Os modelos criados deveriam ser eficientes e precisos, dentro desse contexto, para que não impactem ou limitem o alcance da área de aplicação.

1.1 OBJETIVOS

Como mencionado, o objetivo geral desta dissertação é analisar algoritmos sem referência para avaliação da qualidade de imagens, quando aplicados a imagens de alta resolução.

Para alcançar esse objetivo, foram implementados algoritmos do estado da arte. Os algoritmos que dependem de treinamento foram treinados com diferentes bases de imagens. Avaliamos o desempenho de tais algoritmos nessas bases através de medidas estatísticas, comprovando a eficiência deles. Em seguida, eles foram aplicados a imagens de alta resolução onde avaliamos o desempenho nesse novo contexto. Observamos que os algoritmos apresentados não colocam a resolução como um problema a ser abordado. Ou seja, eles

são propostos para serem aplicados em propósito geral não apenas a imagens de uma determinada resolução.

1.2 ORGANIZAÇÃO DO TRABALHO

Os capítulos restantes desta Dissertação encontram-se estruturados da seguinte forma: o Capítulo **2 (Estado da Arte para Avaliação Sem Referência da Qualidade de Imagens)** descreve com detalhes os principais algoritmos encontrados no estado da arte para o problema de avaliação de qualidade de imagem sem referência, além de introduzir conceitos básicos relacionados à área de visão computacional, e aprendizagem profunda (*deep learning*) necessários para entendimento do trabalho. No Capítulo **3 (Variações Propostas no modelo DIQA)**, são propostas modificações, visando obter melhores resultados no contexto de imagens de alta resolução. O Capítulo **4 (Experimentos)** descreve as medidas e bases de dados de imagens utilizados para avaliar os algoritmos, além de discutir os resultados dos experimentos realizados. No Capítulo **5 (Conclusão)**, são apresentadas as considerações finais sobre os principais tópicos abordados nesta dissertação, incluindo as contribuições alcançadas e as indicações de trabalhos futuros.

2 ESTADO DA ARTE PARA AVALIAÇÃO SEM REFERÊNCIA DA QUALIDADE DE IMAGENS

Neste capítulo, são abordados algoritmos utilizados para avaliação de qualidade de imagem sem referência *No-Reference Image Quality Assessment* (NRIQA). Para melhor compreensão, os algoritmos são divididos em duas categorias: Algoritmos estatísticos e Algoritmo baseado em aprendizagem profunda. Embora muitos métodos de NR-IQA tenham sido propostos nas últimas décadas, neste trabalho, foram selecionados tanto modelos tradicionais, como também modelos publicados recentemente. Os modelos selecionados são detalhados nas seções deste capítulo.

2.1 ALGORITMOS ESTATÍSTICOS

A maioria métodos de avaliação objetiva da qualidade da imagem (IQA) foram propostos para aplicações específicas e são métodos de distorção específica *Distortion Specific* (DS). Na literatura, foram encontrados métodos DS-IQA de última geração, dentre eles podemos citar alguns que foram projetados para medir a nitidez da imagem, degradações JPEG (GOLESTANEH; CHANDLER, 2013), *blockiness artifacts* (LI et al., 2016), distorções de contraste (GU et al., 2016a), *image enhancement* (GU et al., 2014) e *screen content pictures* (GU et al., 2016b). Embora sejam métodos eficientes, a aplicação desses métodos está limitada a um escopo e não pode ser aplicada em cenários mais diversos.

A opção aos métodos DS são os métodos IQA de propósito geral *General Purpose IQA* (GPIQA) que não requerem conhecimento prévio sobre o tipo de distorção na imagem. Em vez de fazer suposições sobre as características de distorções específicas, os métodos GP-IQA fazem suposições sobre as características da imagem (MITTAL et al., 2015). Os métodos são categorizados de acordo com as características que eles extraem da imagem, podendo ser: baseados em *codebooks* (YE et al., 2012), baseados em regressores (TANG; JOSHI; KAPOOR, 2011), ou baseados em Estatísticas de Cena Natural *Natural Scene Statistics* (NSS) (LI et al., 2016). Nos métodos baseados em *codebooks*, o *textitcodebook* é usado para codificar recursos básicos da imagem, como estatísticas de pixel ou a saída de filtros. As abordagens baseadas em NSS assumem que as estatísticas de uma imagem natural são alteradas quando ela é degradada. Por fim, as abordagens baseadas em regressores podem ser treinadas com diferentes características, como NSS, estatísticas de textura, estatísticas de pixel, estatísticas de desfoque ou todas juntas. Nesta dissertação, são abordados apenas os métodos GP, já que o foco é atribuir um escore geral para uma imagem, sem nenhum conhecimento prévio.

2.1.1 Orthogonal Color Planes Patterns - OCPP

O OCPP é um algoritmo baseado em um descritor estatístico proposto por (FREITAS; AKAMINE; FARIAS, 2018), considerando o chamado padrão de planos de cores ortogonais. Este descritor é uma extensão do operador de padrão binário local *Local Binary Pattern* (LBP) (BRAHNAM et al., 2014) que codifica informações espaciais e de cores. Este modelo é inspirado em trabalhos que mostram que o LBP é um descritor de textura eficiente (ZHANG et al., 2014) e (LI; LIN; FANG, 2017). Os descritores LBP podem ser usados com sucesso em métodos IQA porque as deficiências visuais afetam as texturas das imagens e suas estatísticas, ou seja, imagens com qualidades semelhantes têm estatísticas de textura semelhantes. Como o operador LBP é um descritor de textura de baixo custo computacional, os métodos IQA baseados em LBP são geralmente mais rápidos do que outros métodos de última geração que são baseados em recursos de imagem complexos, como (ZHANG et al., 2014) e (LI; LIN; FANG, 2017). Portanto, os métodos IQA baseados em LBP são uma boa abordagem para reduzir a complexidade computacional. O OCPP, por ser baseado em LBP e codificar informações de cores, é mais sensível a distorções de cor e contraste.

2.1.1.1 Descritor OCPP

Como dito, o descritor Orthogonal Color Planes Pattern (OCPP) é uma extensão do operador descritor Local Binary Pattern proposto por (OJALA; PIETIKÄINEN; MÄENPÄÄ, 2000). O descritor LBP pode ser obtido da seguinte forma:

$$LBP_R^P(t_c) = \sum_{p=0}^{P-1} S(t_p - t_c)2^p \quad (2.1)$$

onde R é o raio dos pixels amostrados (que define a região sendo analisada), P é a quantidade de vizinhos e $S(t)$ é a função limite definida por:

$$S(t) = \begin{cases} 1 & t \geq 0, \\ 0 & \text{caso contrário} \end{cases} \quad (2.2)$$

O LBP rotula cada pixel na posição central de acordo com a sua textura, levando em consideração a intensidade do pixel central ($t_c = I(x, y)$) e as intensidades dos pixels vizinhos ($t_p = I(x_p, y_p)$). As coordenadas de t_p são dadas por:

$$x_p = x + R \cos\left(\frac{2\pi * p}{P}\right) \quad (2.3)$$

$$y_p = y - R \sin\left(\frac{2\pi * p}{P}\right) \quad (2.4)$$

A partir dessas equações, podemos ver que o descritor LBP opera em um espaço XY bidimensional, levando em consideração apenas a componente de luminância das imagens.

$$\zeta_{XZ} = LBP_R^{P_{XZ}}(\tau_c) = \sum_{p_{XZ}=0}^{P_{XZ}-1} S(\tau_c - \tau_{XZ})^{p_{XZ}} \quad (2.6)$$

$$\zeta_{YZ} = LBP_R^{P_{YZ}}(\tau_c) = \sum_{p_{YZ}=0}^{P_{YZ}-1} S(\tau_c - \tau_{YZ})^{p_{YZ}} \quad (2.7)$$

Logo, concatenando os LBPs individuais de cada plano, temos o descritor OCPP:

$$OCPP_R^P(\tau_c) = [\zeta_{XY}, \zeta_{XZ}, \zeta_{YZ}]^T \quad (2.8)$$

2.1.1.2 Modelo IQA

O Modelo de avaliação de qualidade de imagens do OCPP é executado após calcular todos os pixels de uma imagem colorida utilizando a Equação 2.8, obtendo um tensor $C(x, y, z)$. Cada coordenada de C corresponde ao rótulo de textura da posição do pixel (x, y) no l -ésimo plano ortogonal, com $x \in [1, M]$, $y \in [1, N]$ e $z \in [1, 3]$, onde M e N são as dimensões da imagem. Em seguida, é feita a extração das informações estatísticas de C , computando seus histogramas.

$$H = \{h(l_1), h(l_2), h(l_3), \dots\} \quad (2.9)$$

onde H representa o histograma e $h(l_i)$ é a frequência do rótulo l_i , que pode ser calculado por:

$$h(l_i) = \sum_{x,y,z} \delta(C(x, y, z), l_i) \quad (2.10)$$

Após o cálculo de H , é realizado um mapeamento dos recursos em um nível de qualidade numérico usando um algoritmo de regressão não linear. O algoritmo de regressão mapeia H em um escore para prever a qualidade de uma imagem. Os autores avaliaram vários tipos de algoritmos de regressão como: *Support Vector Regression* (SVR) (CHANG; LIN, 2011), *Nearest Neighbor Regression* (NNR) (ALTMAN, 1992), e *Random Forest* (RF) (BREIMAN, 2001). Apesar do SVR ser uma técnica amplamente utilizada em métodos NR-IQA (YE et al., 2012) (LIU et al., 2014) (LIU et al., 2014) (MITTAL; MOORTHY; BOVIK, 2012), no trabalho original, os autores relataram que o RF demonstrou melhores resultados em termos de precisão e desempenho.

2.1.2 Blind/Referenceless image spatial quality evaluator (BRISQUE)

O BRISQUE é um algoritmo que utiliza uma estrutura de modelo NSS de coeficientes de luminância normalizados localmente e quantifica a "naturalidade" usando os parâmetros do modelo. Esse algoritmo introduz um novo modelo de estatísticas de produtos de pares de valores de luminância vizinhos. Os parâmetros deste modelo quantificam ainda mais

a naturalidade da imagem e conseqüentemente a qualidade (MITTAL; MOORTHY; BOVIK, 2011).

A abordagem espacial da avaliação do BRISQUE é feita da seguinte maneira: primeiro, devem ser calculadas as luminâncias normalizadas localmente via subtração média local e normalização divisiva (RUDERMAN, 1994). A operação pode ser aplicada a uma imagem de intensidade $I(i, j)$ para produzir:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C} \quad (2.11)$$

onde, $i \in [1, 2, \dots, M]$ e $j \in [1, 2, \dots, N]$ são os índices espaciais, M e N são a altura e a largura da imagem respectivamente, $C = 1$ é uma constante para evitar que existam instabilidades quando o denominador tende a zero. Os valores de $\mu(i, j)$ e $\sigma(i, j)$ podem ser calculados pelas equações 2.12 e 2.13 respectivamente.

$$\mu(i, j) = \sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} I_{k,l}(i, j) \quad (2.12)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} (I_{k,l}(i, j) - \mu(i, j))^2} \quad (2.13)$$

onde $w = \{w_{k,l} \mid k \in [-K, \dots, K], l \in [-L, \dots, L]\}$ é uma função de ponderação gaussiana circularmente simétrica, amostrada para três desvios padrão e redimensionada para unidade de volume. A versão publicada pelos autores e utilizada nessa dissertação utiliza os valores de $K = L = 3$. O pré-processamento é calculado de acordo com a Equação 2.11 e refere-se às luminâncias transformadas $I(i, j)$ como coeficientes da Média Subtraída do Contraste Normalizado (MSCN). Uma Distribuição Gaussiana Generalizada (GGD) é usada para capturar efetivamente um espectro mais amplo das estatísticas de imagens distorcidas, que normalmente exibem mudanças no comportamento da cauda das distribuições de coeficientes empíricos (SHARIFI; LEON-GARCIA, 1995). A GGD com média zero é dada por:

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp\left(-\left(\frac{|x|}{\beta}\right)^\alpha\right) \quad (2.14)$$

onde

$$\beta = \alpha \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}} \quad (2.15)$$

e

$$\Gamma(a) = \int_0^\infty t^{(a-1)} e^{-t} dt \quad a > 0 \quad (2.16)$$

O parâmetro α controla a “forma” da distribuição enquanto σ^2 controla a variância. É utilizada a distribuição de média zero, tendo em vista que a distribuições dos coeficientes

das Médias Subtraídas do Contraste Normalizado (MSCN) são simétricas. Os parâmetros do GGD (α, σ^2) são estimados usando a abordagem baseada no *moment-matching* proposta por (SHARIFI; LEON-GARCIA, 1995).

Um modelo paramétrico proposto por (MITTAL; MOORTHY; BOVIK, 2012) é utilizado para ajustar as distribuições empíricas MSCN de imagens distorcidas e não distorcidas. Cada imagem tem dois parâmetros (α, σ^2) estimados, considerando um ajuste GGD dos coeficientes MSCN. Esses valores formam o primeiro conjunto características usados para avaliação da imagem.

Para as demais características, foram modelados as relações entre pixel vizinhos. Os coeficientes MSCN são homogêneos para imagens “puras”, já os sinais dos coeficientes adjacentes também exibem uma estrutura regular, que é perturbada na presença de distorção. Dessa forma, a estrutura das relações estatísticas entre pixels vizinhos utiliza as distribuições empíricas de produtos de pares de coeficientes MSCN vizinhos ao longo de quatro orientações: (H) Horizontal, (V) Vertical, (D1) Diagonal principal e (D2) Diagonal secundária. Elas são dadas por:

$$H(i, j) = \hat{I}(i, j)\hat{I}(i, j + 1) \quad (2.17)$$

$$V(i, j) = \hat{I}(i, j)\hat{I}(i + 1, j) \quad (2.18)$$

$$D1(i, j) = \hat{I}(i, j)\hat{I}(i + 1, j + 1) \quad (2.19)$$

$$D2(i, j) = \hat{I}(i, j)\hat{I}(i + 1, j - 1) \quad (2.20)$$

para, $i \in [1, 2, \dots, M]$ e $j \in [1, 2, \dots, N]$.

Com o modelo de coeficiente gaussiano, e assumindo que os coeficiente MSCN são de média zero e variância unitária, na ausência de distorção, esses produtos obedecem à seguinte distribuição:

$$f(x, \rho) = \frac{\exp\left(\frac{|x|\rho}{1-\rho^2}\right)K_0\left(\frac{|x|}{1-\rho^2}\right)}{\pi\sqrt{(1-\rho^2)}} \quad (2.21)$$

onde, f é uma função de densidade de probabilidade assimétrica, ρ denota o coeficiente de correlação de coeficientes adjacentes e K_0 é a função de Bessel modificada de segundo tipo. A Distribuição Gaussiana Generalizada Assimétrica (AGGD) é usada como alternativa prática à $f(x, \rho)$ que, apesar de ser um bom modelo dos histogramas empíricos de produtos de coeficientes normalizados adjacentes, possui apenas um único parâmetro e, sendo assim, não fornece um bom ajuste aos histogramas empíricos de produtos de coeficientes de imagens distorcidas como mostrado por (LASMAR; STITOU; BERTHOUMIEU, 2009).

O AGGD modelo zero pode ser obtida pela seguinte equação:

$$f(x; \nu, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{\nu}{(\beta_l + \beta_r)\Gamma(\frac{1}{\nu})} \exp(-(\frac{-x}{\beta_l})^\nu) & x \leq 0, \\ \frac{\nu}{(\beta_l + \beta_r)\Gamma(\frac{1}{\nu})} \exp(-(\frac{x}{\beta_r})^\nu) & x \geq 0 \end{cases} \quad (2.22)$$

onde

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}} \quad (2.23)$$

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma(\frac{1}{\nu})}{\Gamma(\frac{3}{\nu})}} \quad (2.24)$$

O parâmetro ν controla a “forma” da distribuição enquanto, σ_l^2 e σ_r^2 são parâmetros de escala que controlam a propagação em cada lado do modo, respectivamente. O AGGD generaliza ainda mais a distribuição gaussiana generalizada (GGD) (SHARIFI; LEON-GARCIA, 1995), permitindo assimetria na distribuição. A inclinação da distribuição é uma função de parâmetros de escala esquerdo e direito. Se $\sigma_l^2 = \sigma_r^2$, então o AGGD se reduz ao GGD. Os parâmetros do AGGD ($\nu, \sigma_l^2, \sigma_r^2$) são estimados usando a abordagem baseada no *moment-matching* proposta por (LASMAR; STITOU; BERTHOUMIEU, 2009).

Os parâmetros ($\eta, \nu, \sigma_l^2, \sigma_r^2$) que melhor se ajustam para a AGGD são extraídos quando η é dado por:

$$\eta = (\beta_r - \beta_l) \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})} \quad (2.25)$$

Assim, para cada produto emparelhado, são calculados 16 parâmetros (4 parâmetros por orientação \times 4 orientações), gerando o segundo conjunto de características.

A Tabela 1 resume todas as características extraídas pelo BRISQUE.

Tabela 1 – Resumo das características extraídas pelo BRISQUE

Característica	Descrição	Procedimento de cálculo
$f_1 - f_2$	forma e variância	GGD para os coeficientes MSCN
$f_3 - f_6$	forma, média, variância esquerda, variância direita	AGGD para o par H
$f_7 - f_{10}$	forma, média, variância esquerda, variância direita	AGGD para o par V
$f_{11} - f_{14}$	forma, média, variância esquerda, variância direita	AGGD para o par D1
$f_{15} - f_{18}$	forma, média, variância esquerda, variância direita	AGGD para o par D2

Fonte: Adaptado de (MITTAL; MOORTHY; BOVIK, 2012)

A avaliação de qualidade é realizada por meio de um mapeamento aprendido a partir das características espaciais para pontuações de qualidade usando um módulo de regressão, produzindo uma medida da qualidade da imagem. A estrutura é genérica o suficiente para permitir o uso de qualquer regressor. Um regressor (SVR) de máquina de vetor de suporte (SVM) (SCHÖLKOPF et al., 2000) é usado; o SVR é geralmente conhecido por ser capaz de lidar com dados de alta dimensão (BURGES, 1998).

O algoritmo quantifica apenas a “naturalidade” (ou falta dela) na imagem devido à presença de distorções. BRISQUE é altamente competitivo com todos os algoritmos NR-IQA. Mostrou ser computacionalmente eficiente, tornando-se uma boa opção em termos de precisão e processamento.

2.2 ALGORITMO BASEADO EM APRENDIZAGEM PROFUNDA

As redes neurais convolucionais (CNN) formam o modelo de aprendizado profundo mais popular atualmente devido à sua forte capacidade de representação e desempenho impressionante. As CNNs têm sido aplicadas com sucesso a vários problemas de visão computacional e processamento de imagens. Os algoritmos baseados em *Deep Learning* derivam das Redes Neurais Convolucionais (CNN) (LECUN; BENGIO, 1998) que têm sido uma das propostas mais utilizadas, na última década, para uma variedade de problemas relacionados a reconhecimento de padrões. As CNN são consideradas o estado da arte em problemas relacionados a visão computacional como detecção, segmentação, reconhecimento e avaliação. Apesar das redes neurais tradicionais já existirem há algumas décadas, as redes convolucionais se tornaram mais difundidas, em anos recentes, com o avanço no poder de processamento adquirido por meio do uso das GPUs para o processamento paralelo dos modelos.

O desempenho das redes neurais *deep* depende muito do número de dados de treinamento (desconsiderando situações de aproveitamento de arquiteturas já treinadas com o uso de *transfer learning*). No entanto, as bases de dados de imagens públicas, atualmente disponíveis, são muito menores em comparação com o conjunto de dados de visão computacional típico para aprendizado profundo. Por exemplo, a base de dados LIVE IQA (SHEIKH et al., 2005), usada para métodos de avaliação automática de qualidade de imagem, contém de 174 a 233 imagens para cada tipo de distorção representada, enquanto o conjunto de dados amplamente utilizado para reconhecimento de imagens (ImageNet) contém mais de 1,2 milhão de dados rotulados (RUSSAKOVSKY et al., 2015). Além disso, obter rótulos subjetivos confiáveis em larga escala é muito difícil, impreciso e custoso. Ao contrário dos rótulos de classificação, a construção de um banco de dados IQA requer um experimento psicométrico complexo e demorado (Hosu et al., 2020). Além disso, o processo perceptivo do sistema visual humano inclui vários processos complexos e subjetivos, o que dificulta ainda mais o treinamento de um modelo *deep* com um conjunto de dados limitado.

Das propostas existentes na literatura, o DIQA (Deep Image Quality Assessment) (KIM; NGUYEN; LEE, 2018) foi escolhido considerando tempo de treinamento nas bases adotadas e tempo de avaliação de uma imagem nos testes com as imagens de alta resolução. Os modelos Koncept512 e o Koncept1024 propostos por (Hosu et al., 2020) também foram avaliados, porém não apresentaram resultados promissores nos primeiros experimentos e não serão considerados nesse trabalho. Nos experimentos, o DIQA, mesmo com uma

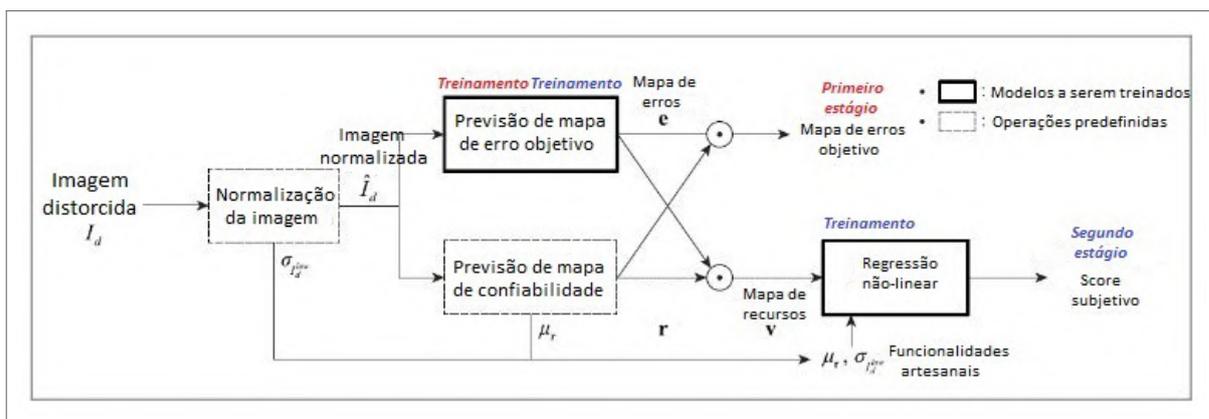
arquitetura com poucas camadas, alcançou o limite da capacidade dos computadores disponíveis com imagens de 12.000 x 9.000 pixels.

2.2.1 Deep Image Quality Assessment - DIQA

Proposto em (KIM; NGUYEN; LEE, 2018), o *Deep Image Quality Assessment* (DIQA) é um modelo baseado em CNN para avaliação de qualidade de imagens sem referência (NR-IQA). Por ser um modelo de inteligência artificial, espera-se que o escore atribuído a uma imagem considere aspectos objetivos e subjetivos, modelando da melhor forma o comportamento do sistema visual humano.

Esse modelo é dividido em várias etapas e é detalhado nesta seção; o diagrama de fluxo pode ser observado na Figura 4. A imagem de entrada é normalizada e passa por dois caminhos: (1) uma ramificação CNN e (2) uma ramificação de previsão do mapa de confiabilidade. Na primeira etapa de treinamento, o ramo CNN é treinado para prever um mapa de erro objetivo. O mapa de erro de *ground-truth* e_{gt} é obtido, comparando as imagens de referência e distorcidas. Na segunda etapa, o modelo é treinado utilizando imagens públicas rotuladas com uma pontuação de qualidade atribuídas por humanos. E como saída, temos o score S , que é a nota de qualidade atribuída a uma imagem. S é chamado de score subjetivo pois, é esperado que a rede aprenda os conceitos subjetivos utilizados para avaliar uma imagem e que os mesmos sejam utilizados para prever a pontuação. Em cada estágio, o mapa de confiabilidade r é complementado para compensar a imprecisão em regiões homogêneas. Neste trabalho, vamos utilizar apenas o segundo estágio do DIQA, que é onde podemos obter a pontuação de qualidade de uma imagem.

Figura 4 – Fluxo fim-a-fim do DIQA



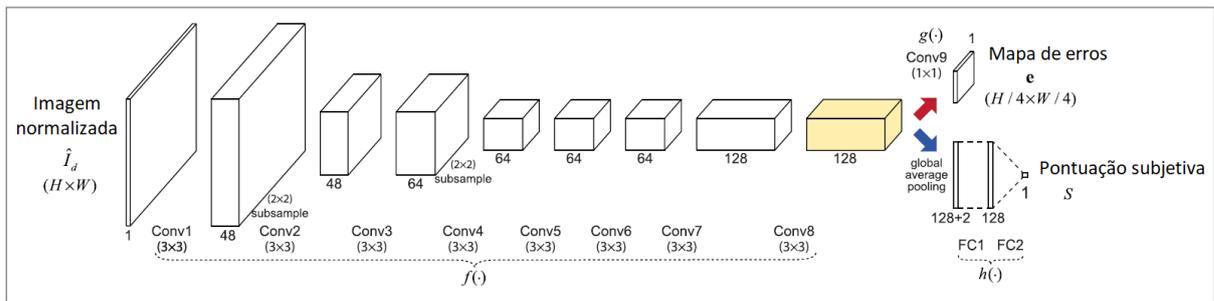
Fonte: Adaptado de (KIM; NGUYEN; LEE, 2018)

2.2.1.1 Arquitetura do Modelo

O projeto da arquitetura CNN proposta é motivado por (SIMONYAN; ZISSERMAN, 2014). A estrutura do DIQA é mostrada na Figura 5. No que se refere à parte de previsão do mapa

de erro, o modelo consiste em camadas convolucionais, onde as bordas são preenchidas com zeros antes de cada convolução. Dessa forma, a saída não perde a informação relativa da posição do pixel. Cada camada, exceto a última, possui um filtro 3x3 e uma Unidade Linear Retificada (RELU) (NAIR; HINTON, 2010). Para a operação de *downsampling*, é usada uma convolução com *stride* de 2. A saída de Conv8 (ver Figura 5) é chamada como um mapa de características (preenchido em amarelo na figura), que é re-utilizado para o segundo estágio de treinamento. Nesse estágio, o mapa de características extraído é alimentado na camada de *pooling* de média global (Global Average Pooling (GAP)), seguida por duas camadas totalmente conectadas (FC1 e FC2, respectivamente). Além disso, são usados dois recursos manuais, que são explicados mais adiante. Os recursos artesanais são concatenados com os recursos agrupados antes do FC1 e, em seguida, regredidos em uma pontuação subjetiva. Por conveniência, é denotado o procedimento de Conv1 a Conv8 por $f(\cdot)$ e o procedimento incluindo FC1 e FC2 por $h(\cdot)$.

Figura 5 – Arquitetura DIQA para previsão de qualidade de imagem



Fonte: Adaptado de (KIM; NGUYEN; LEE, 2018)

2.2.1.2 Normalização da Imagem

Como pré-processamento, as imagens de entrada são primeiro convertidas em escala de cinza e são subtraídas da própria imagem filtrada por um filtro passa-baixa. Seja I_r uma imagem de referência e I_d a imagem distorcida correspondente. As versões normalizadas são então denotadas por \hat{I}_r e \hat{I}_d , respectivamente. A imagem de baixa frequência é obtida, reduzindo a imagem de entrada para 1/4 e aumentando-a novamente para o tamanho original, que é denotado por I_r^{low} e I_d^{low} . Um filtro passa-baixa Gaussiano e subamostragem são usados para redimensionar as imagens. A Equação 2.26 descreve a operação utilizada no processo de normalização da imagem.

$$\hat{I} = I_{gray} - I^{low} \quad (2.26)$$

onde I_{gray} é a imagem em escala de cinza e I^{low} é a imagem filtrada por um filtro passa-baixa. O filtro gaussiano aplicado foi configurado com os seguintes parâmetros:

- Tamanho do kernel: 16

- Sigma: 7/6

As distorções pouco afetam os componentes de baixa frequência das imagens. Por exemplo, ruído gaussiano branco adiciona componentes aleatórios de alta frequência às imagens e o borramento gaussiano (GB) remove detalhes de alta frequência. As distorções devido ao JPEG e JPEG2000 (JP2K) podem ser modeladas por uma combinação desses artefatos (SHEIKH; BOVIK, 2006) e o SVH não é sensível a mudanças em componentes de baixa frequência. Embora existam pequenas distorções na banda de baixa frequência, o SVH dificilmente as percebe. Apesar de haver vantagens em empregar esse esquema de normalização, há também a desvantagem de perder informações. Para compensar isso, dois recursos artesanais são complementados no segundo estágio de treinamento.

2.2.1.3 Avaliação subjetiva

Para conseguir prever pontuações subjetivas, o DIQA precisa ser treinado. A sub-rede $f(\cdot)$ é re-utilizada e conectada com uma camada de *pooling* de média global, seguida pelas camadas totalmente conectadas, conforme mostrado na Figura 5. O mapa de recursos é calculado sobre o domínio espacial, levando a um vetor de recursos de 128-D. Para compensar a perda de informação, são consideradas duas características artesanais adicionais: a média do mapa de confiabilidade não normalizado μ_r e o desvio padrão da versão de baixa frequência da imagem distorcida $\sigma_{I_d}^{low}$. Se a imagem distorcida estiver muito borrada, a área confiável ficará muito pequena. Nesse caso, a textura geral da imagem distorcida se torna uma característica importante, que pode ser capturada por μ_r . Portanto, a função de perda (*loss*) é definida como:

$$L_2(I_d; \theta_f, \theta_h) = \|(h(v, \mu_r, \sigma_{I_d}^{low}) - S)\|_2^2 \quad (2.27)$$

onde $f(\cdot)$ é uma função de regressão não linear, S é a pontuação subjetiva de base da imagem distorcida de entrada e v é o vetor de características agrupado. v é definido por:

$$v = GAP(f(\hat{I}_d; \theta_f)) \quad (2.28)$$

onde GAP indica a operação *Global Average Pooling*.

2.2.1.4 Treinamento

Para melhor convergência, o Otimizador de Estimação de Momento Adaptativo (ADAM) (KINGMA; BA, 2014) com Nesterov *momentum* (DOZAT, 2016) foram utilizados para alterar o método de gradiente descendente estocástico regular. Os hiper-parâmetros utilizados pelos autores no trabalho original, são valores padrão sugeridos na literatura por (KINGMA; BA, 2014). Esses valores são usados para o ADAM, e o parâmetro *momentum* é ajustado para 0,9. A taxa de aprendizado no trabalho original foi definida como: 2×10^{-4} ,

como sugerido pela literatura. Além disso, durante o treinamento, as taxas de aprendizado para as camadas pré-treinadas são multiplicadas por 0,1. Para o decaimento de peso, a regularização L2 é aplicada a todas as camadas (penalidade L2 multiplicada por 2×10^{-4}).

Para treinar o DIQA, usando imagens de vários tamanhos, como no banco de dados LIVE IQA (SHEIKH; SABIR; BOVIK, 2006), cada imagem de entrada deve ser dividida em vários *patches* de mesmo tamanho. Aqui, o passo da janela deslizante é determinado pelo tamanho do *patch* e o número de pixels ignorados ao redor das bordas para evitar regiões sobrepostas. Durante o treinamento do segundo estágio, todos os trechos que compõem uma imagem devem estar nos mesmos mini lotes (KIM; LEE, 2016).

O DIQA apresentou bons resultados quando comparado com outros métodos de NR-IQA. O modelo se mostrou eficiente utilizando as bases de dados de imagens públicas, gerando pontuações de qualidade muito próximas às pontuações atribuídas por humanos. Podemos ver alguns desses testes no Capítulo 4.

3 VARIAÇÕES PROPOSTAS NO MODELO DIQA

De todos os modelos apresentados no capítulo anterior, o DIQA mostrou-se o mais robusto, tanto em relação à coerência dos escores de saída quanto em relação ao tempo de avaliação das imagens. Dito isso, este trabalho propõe alterações no DIQA para tentar aprimorar o modelo para o cenário de imagens de alta resolução de dispositivos móveis.

3.1 PROBLEMA

A maioria dos modelos propostos na literatura são modelos treinados e avaliados em bases públicas e que nem sempre modelam as imagens digitais encontradas em situações reais. Isso por causa do alto custo envolvido na criação de bases de dados, considerando que, para cada imagem, é necessário a avaliação subjetiva de várias pessoas e em condições específicas.

Nas bases investigadas, as imagens possuem baixa resolução, quando comparadas a imagens reais geradas por dispositivos móveis modernos. A maior resolução encontrada nessas bases públicas foi de 1024x768 pixels. Dispositivos móveis de última geração capturam imagens 5 ou 6 vezes maiores que essas bases o que justifica um desempenho não tão bom quando aplicamos os modelos nessas imagens. A diminuição da resolução da imagem pode esconder ruídos, aparentando tornar a imagem reduzida de melhor qualidade do que ela realmente tinha.

3.2 PROPOSTAS DE APRIMORAMENTO DO DIQA

Para resolver o problema da aplicação do DIQA em imagens de alta resolução sem criação de uma nova base, propomos algumas modificações no modelo DIQA, conforme apresentado as próximas seções.

3.2.1 Normalização (pré-processamento)

No DIQA, existe uma etapa chamada de Normalização da Imagem, como pode ser visto na Figura 4, na qual é feito um pré-processamento da imagem de entrada. Nessa etapa de pré-processamento, são aplicados filtros para remover ruídos que não são perceptíveis para o Sistema Visual Humano (SVH). Os filtros, no entanto, foram dimensionados para imagens de baixa resolução, como as imagens de treinamento, não tendo o mesmo efeito em imagens de alta resolução. Como a normalização da imagem não tem o efeito visual esperado para as imagens de alta resolução, a execução das etapas seguintes do modelo são comprometidas, tornando a avaliação final menos precisa. Afinal, as imagens não estão sendo avaliadas nas mesmas condições das imagens usadas na etapa de treinamento.

A Figura 4 ilustra todas as etapas do modelo DIQA. A proposta é remover a etapa de normalização de imagens para que esses filtros sejam aprendidos pela própria rede de aprendizagem profunda de acordo com as imagens de entrada.

Um estudo feito pelo próprio autor do DIQA mostra que existe uma pequena variação entre usar ou não a etapa de normalização da imagem. Dessa forma, propomos a remoção do procedimento durante as etapas de treinamento e avaliação do modelo. Supomos que esse impacto seja muito maior quando se utiliza imagens de alta resolução.

3.2.2 Super-resolução

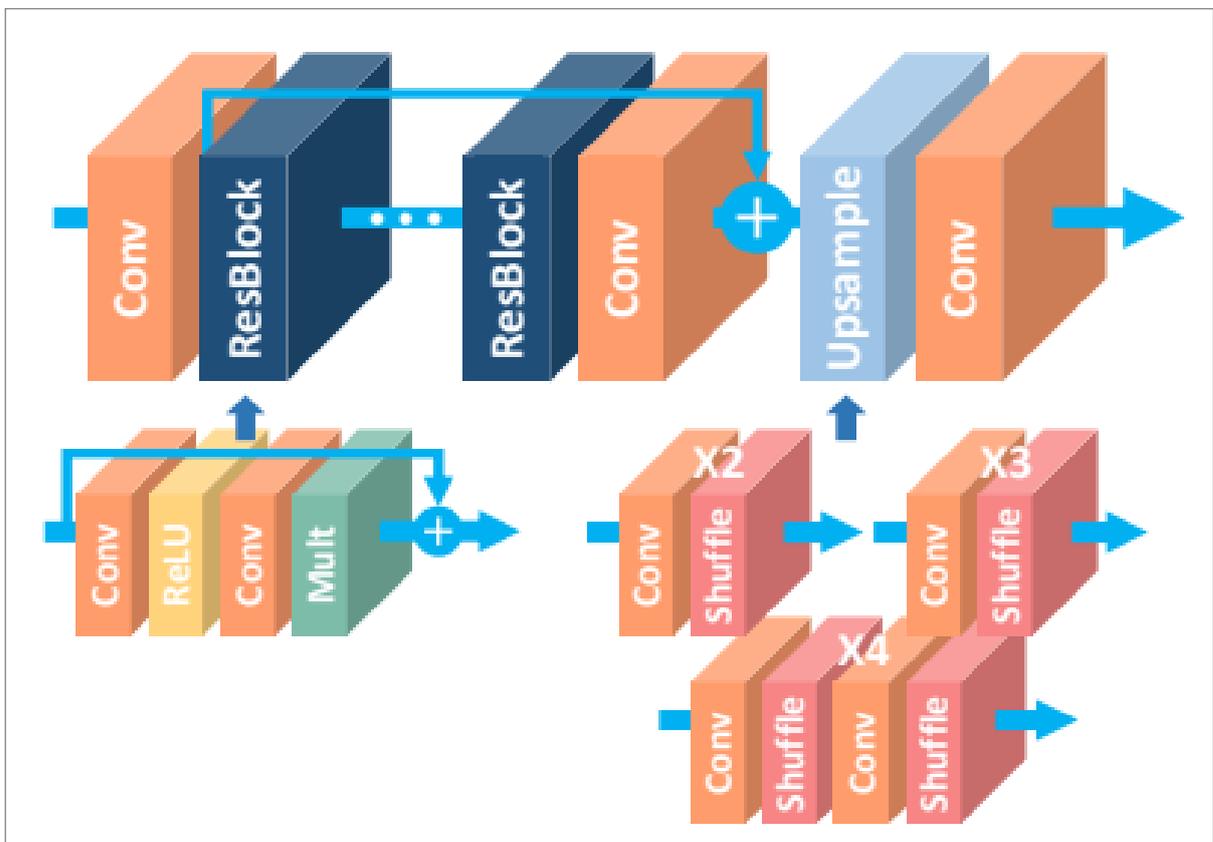
Como citado anteriormente, o custo de se criar uma base de imagens rotuladas é muito alto, requer muito esforço, minucioso e exaustivo, e envolve diversas pessoas (Hosu et al., 2020). Dito isso, propomos criar uma base de dados artificial baseada na LIVE IQA, porém com imagens em alta resolução. Para tal, utilizamos algoritmos de aprendizagem profunda para aumentar a resolução das imagens em 5 ou 6 vezes. Com isso, as imagens da base de treinamento ficam com resolução similar às imagens sob análise. Porém, tudo isso deve ser feito sem perder as características originais das imagens para que os escores sejam preservados.

A rede *deep* utilizada para criação da base de dados artificial foi a *Enhanced Deep Super Resolution Network* (EDSR) (LIM et al., 2017), um modelo robusto que foi o vencedor do "New Trends in Image Restoration and Enhancement workshop and challenge on image super-resolution" (NTIRE) de 2017. A arquitetura do modelo EDSR pode ser visualizada na Figura 6.

O modelo EDSR é um modelo baseado na ResNet (He et al., 2016), no qual, foram removidas as camadas de *Batch Normalization* (BN) e a função de ativação final ReLU, dos blocos residuais, como mostrado na Figura 7.

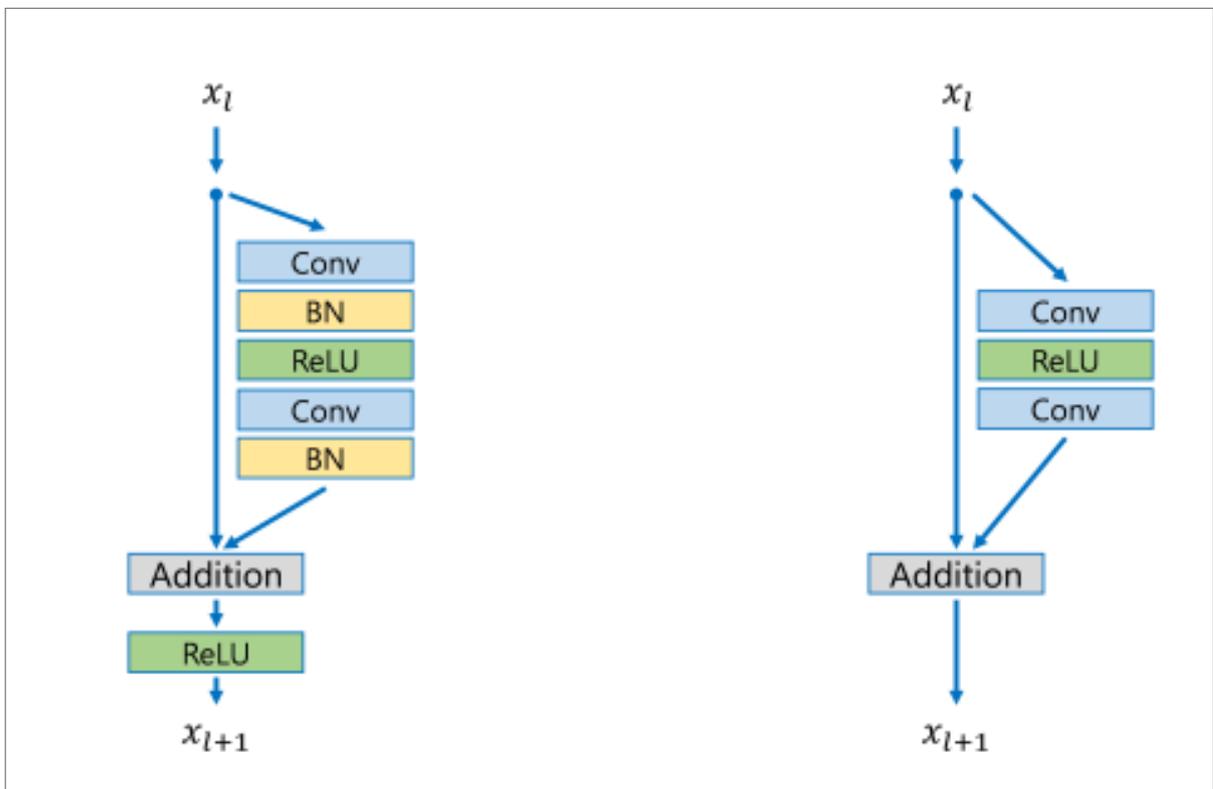
Nesta dissertação, foi utilizado o modelo EDSR pré-treinado disponibilizado pelos autores. Esse modelo foi aplicado nas imagens da base LIVE IQA, aumentando a resolução das imagens em 4x, gerando um banco de imagens artificial: LIVE IQA Super Resolution (LIVE IQA SR). A base LIVE IQA SR possui resolução muito próxima a das imagens capturadas por dispositivos móveis da atualidade. Assim, treinamos o modelo DIQA utilizando a base de imagens LIVE IQA SR. Como mostrado por (LIM et al., 2017) a qualidade da imagem é preservada, com isso, o seu score (MOS) não precisa ser ajustado. Assim, utilizamos os mesmos valores de MOS da base LIVE IQA original. A abordagem descrita na seção anterior também será aplicada, já que a etapa de normalização da imagem também está presente nas imagens de treinamento. Dessa forma, essa proposta não utilizará a etapa de normalização do DIQA.

Figura 6 – Arquitetura do modelo EDSR



Fonte: (LIM et al., 2017)

Figura 7 – Blocos residuais da ResNet (esquerda) e EDSR (direita)



Fonte: Adaptado de (LIM et al., 2017)

4 EXPERIMENTOS

Este capítulo tem como objetivo apresentar como foram elaborados os experimentos para avaliar os métodos propostos no Capítulo 3 em comparação ao modelo base *DIQA*. Na Seção 4.1 são apresentadas as medidas que foram utilizadas para avaliar os modelos e como elas funcionam. A Seção 4.2 introduz as bases de dados de imagens utilizadas nos experimentos. A Seção 4.3 detalha o ambiente que foi utilizado para executar todos os experimentos, incluindo o treinamento e avaliação. E na Seção 4.5 são apresentados os resultados, com a avaliação e comparação dos modelos.

4.1 MEDIDAS PARA AVALIAÇÃO

Esta seção apresenta as medidas que são utilizadas para a avaliação e comparação dos modelos. Semelhante aos trabalhos da área, são usadas as medidas: *Spearman Rank Order Correlacionados Coefficient* (SROCC), *Pearson Linear Correlation Coefficient* (PLCC), *Kendall Rank Order Correlation Coefficient* (KROCC) e *Root Mean Square Error* (RMSE). Quanto maior o valor de SROCC, KROCC e PLCC, e quanto menor o valor de RMSE, melhor o desempenho do modelo. Nas próximas subseções são detalhadas cada uma dessas medidas.

4.1.1 Pearson Linear Correlation Coefficient - PLCC

O coeficiente de correlação linear de Pearson (PLCC) define a correlação linear entre dois conjuntos de dados. É uma medida normalizada da covariância, com resultado entre -1 e 1. O PLCC é usado como medida da precisão, para mostrar o quão próximo estão os *scores* previstos e os *scores* reais (BULL, 2014). A fórmula utilizada para calcular esse coeficiente para um conjunto de N pares é descrita na Equação 4.1

$$PLCC = \frac{\sum_{i=0}^{N-1} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{N-1} (x_i - \bar{x})^2} \sqrt{\sum_{i=0}^{N-1} (y_i - \bar{y})^2}} \quad (4.1)$$

onde x_i representa o escore real e y_i o escore de previsto para um par de amostras i .

4.1.2 Spearman Rank Order Correlation Coefficient - SROCC

O coeficiente de correlação de classificação de Spearman mede a força e a direção da associação entre duas variáveis. Ela basicamente fornece a medida da monotonicidade da relação entre duas variáveis, ou seja, quão bem a relação entre duas variáveis pode ser representada usando uma função monotônica. O resultado da SROCC é um valor entre -1 e 1. O valor 1, significa uma associação perfeita. O valor 0, significa que não existe associação e o valor -1 significa uma associação negativa perfeita. Convencionalmente, o

coeficiente de correlação de ordem de classificação de Spearman é usado para medir o grau em que as previsões do modelo se correlacionam com as magnitudes relativas dos escores subjetivos. O cálculo do coeficiente pode ser realizado utilizando a seguinte fórmula:

$$SROC = \frac{\sum_{i=0}^{N-1} (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=0}^{N-1} (X_i - \bar{X})^2} \sqrt{\sum_{i=0}^{N-1} (Y_i - \bar{Y})^2}} \quad (4.2)$$

onde x_i e y_i tem a mesma interpretação que na Equação 4.1, X_i é a ordem de classificação de x_i , Y_i é a ordem de classificação de y_i e \bar{X} e \bar{Y} são seus valores medianos. (BULL, 2014).

4.1.3 Kendall's Rank Order Correlation Coefficient - KROCC

KROCC é um teste não paramétrico que mede a força da dependência entre duas variáveis. Para essa medida, os valores próximos a 1 indicam forte concordância e valores próximos a -1 indicam forte discordância. A equação 4.3 é a fórmula usada para calcular o valor do coeficiente de correlação de classificação de Kendall (KENDALL, 1945).

$$KROCC = \frac{(P - Q)}{\sqrt{(P + Q + T) * (P + Q + U)}} \quad (4.3)$$

onde P é o número de pares concordantes, Q o número de pares discordantes, T o número de empates apenas em x e U o número de empates apenas em y . Se ocorrer um empate para o mesmo par em x e y , ele não é adicionado a T ou U .

4.1.4 Root Mean Square Error - RMSE

O RMSE é uma medida de risco que nos ajuda a determinar a diferença quadrática média entre o valor previsto e o valor real do escore. Essa é uma medida de visa identificar quão longe ou espalhados estão as previsões da linha de regressão. Como visto na equação 4.4 o valor pode ser calculado pela raiz quadrada das diferenças quadradas entres o valor estimado e o valor real do escore (CHAI; DRAXLER, 2014).

$$RMSE = \sqrt{\frac{\sum_{i=0}^{N-1} (x_i - y_i)^2}{N}} \quad (4.4)$$

onde N é a quantidade de pares de amostras, x_i representa o escore real e y_i o escore previsto para um par de amostras i .

4.2 BASES

Objetivando avaliar a acurácia dos modelos propostos e compará-los com os modelos da literatura, foram executados experimentos variando a base de dados de imagens de treinamento. Nesta seção, são detalhadas as três bases de imagens utilizadas: LIVE IQA, LIVE In the Wild e a KonIQ-10k.

4.2.1 Base LIVE IQA

Uma base de dados de imagens disponibilizada pelo [Laboratório de Engenharia de Imagem e Vídeo \(LIVE\)](#) da Universidade do Texas, é uma base de dados tradicional e popular, amplamente utilizada em outros estudos da literatura. O banco de dados de avaliação de qualidade de imagem LIVE (LIVE IQA) contém 779 imagens separadas pelo tipo de distorção e com classificações MOS (*Mean OPinion Score*). As distorções da versão 2, utilizada nesta Dissertação, incluem JPEG (169 imagens), JPEG2000 (175 imagens), ruído branco (145 imagens), borramento gaussiano (145 imagens) e JPEG2000 com erros de bits no canal de desvanecimento Rayleigh simulado (145 imagens) ([SHEIKH et al., 2005](#)). As imagens dessa base de dados possuem diversas dimensões diferentes (em pixels): 505x632, 512x768, 505x634, 720x480, 512x640, 453x618, 438x634, 488x610, e 482x627.

4.2.2 Base LIVE In the Wild

Esta base é derivada da base LIVE IQA, porém com imagens naturais, autenticamente distorcidas, sem introduzir artificialmente quaisquer efeitos além daqueles que ocorreram durante a captura, processamento e armazenamento pelo dispositivo de um usuário. O banco de dados da *Live In the Wild* contém 1.162 imagens. As avaliações das imagens foram obtidas a partir de um estudo subjetivo online aproveitando o sistema de *crowdsourcing* Amazon Mechanical Turk. Cada imagem da base de dados foi visualizada e avaliada online em uma escala de qualidade contínua por uma média de 175 assuntos exclusivos. A média e a variância do [MOS](#) obtidos nas avaliações subjetivas, juntamente com as imagens, estão disponíveis como parte do banco de dados ([GHADIYARAM; BOVIK, 2015a](#)). A base possui imagens com dimensões (em pixels): 960x640, 608x608, e 500x500.

Apesar da falta de controle sobre os ambientes experimentais dos numerosos participantes do estudo, os resultados reportados no artigo ([GHADIYARAM; BOVIK, 2015b](#)) mostram uma excelente consistência interna das pontuações subjetivas.

4.2.3 KonIQ-10k

Esta é um dos maiores conjuntos de dados para avaliação de qualidade de imagem, contendo 10.073 imagens rotuladas com pontuação de qualidade (MOS). É um banco de dados *in-the-wild* que retrata, de forma autêntica, as distorções e diversidade de conteúdo encontradas naturalmente. Para rotular essas imagens, também foi utilizado o modelo de *crowdsourcing*, obtendo, assim, 1,2 milhão de classificações de qualidade confiáveis de 1.459 pessoas, abrindo caminho para modelos IQA generalistas ([Hosu et al., 2020](#)). Todas as imagens dessa base de dados possuem as mesmas dimensões: 384x512 pixels.

4.3 AMBIENTE

Para realização dos experimentos utilizando os modelos encontrados na literatura e as variações propostas no Capítulo 3, foi utilizada uma máquina nas configurações descritas na Tabela 2. Alguns modelos como o BRISQUE e o OCPP não exploram recursos de GPU. Então, para deixar a comparação justa, a GPU foi utilizada apenas nas etapas de treinamento dos modelos. Na etapa de avaliação da qualidade das imagens, utilizamos apenas a CPU. Para os métodos que se utilizam de algum número aleatório, como a inicialização dos pesos dos modelos, foi definida uma $seed = 1111$ para manter a consistência dos experimentos e que seja possível reproduzi-los posteriormente. *Seed* é uma variável definida para que os valores aleatórios sejam os mesmos. Ela é utilizada na definição dos otimizadores, para gerar os pesos iniciais aleatoriamente. Com o mesmo *seed*, qualquer um pode gerar os mesmos pesos iniciais. Isso deixa o algoritmo mais determinístico e replicável.

Tabela 2 – Configurações de hardware utilizadas na execução dos experimentos

Sistema Operacional	Ubuntu 20.04 LTS
CUDA	CUDA 10.2
GPU	Tesla T4, 2560 CUDA cores, 15GB GDDR5 VRAM
Processador	Xeon @2.0Ghz, 1 Core, 2 Threads
Memória	13GB disponível
Armazenamento	40 GB disponível
Biblioteca	Tensorflow 2

Fonte: Elaborado pelo autor

4.4 EXPERIMENTOS COM O DIQA

Antes de apresentar os resultados das variações propostas nas diferentes bases, vamos explicar os experimentos que levaram à escolha do DIQA como o representante da classe de algoritmos baseada em *deep learning*. Os testes foram feitos com as bases LIVE e KonIQ-10K, com os resultados avaliados de acordo com as medidas apresentadas na Seção 4.1 (PLCC, SROC, RMSE e KROCC). A expectativa é que bons algoritmos têm PLCC, SROCC e KROCC com valores próximos de 1 e RMSE o mais baixo possível (próximo de zero, já que é uma medida de erro).

A Tabela 3 apresenta os resultados das medidas quando o DIQA é treinado nas bases LIVE e KonIQ-10K e testado em imagens dessas mesmas bases.

Dessas bases, selecionamos três imagens de cada para ilustrar a predição do método em comparação com o MOS definido por avaliadores humanos. Esses resultados podem ser vistos nas Tabelas 4 e 5.

Tabela 3 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência quando treinado e aplicado nas bases LIVE e KonIQ-10K.

Base	SROCC \uparrow	KROCC \uparrow	PLCC \uparrow	RMSE \downarrow
LIVE IQA	0,9808	0,8803	0,9857	0,0487
KonIQ-10K	0,9508	0,8123	0,9631	0,0477

Fonte: Elaborado pelo autor

Tabela 4 – Escores preditos pelo algoritmo DIQA em comparação os escores definidos por avaliadores humanos: exemplos da base LIVE IQA.

Imagem	Escore predito	Escore humano
	0,7535	0,7406
	0,4867	0,4627
	0,5883	0,6715

Fonte: Elaborado pelo autor

Na Tabela 4, podemos observar imagens variadas. Por exemplo, na imagem das araras da base LIVE-IQA, o fundo da imagem está borrado o que implica em uma esperada queda de escore, como realmente acontece. A imagem da mulher de chapéu tem uma saturação do tom vermelho, novamente gerando uma expectativa de diminuição de nota de avaliação. Já as imagens da Tabela 5 não aparentam distorções de imagem ou de cor perceptíveis. Em ambas as tabelas, os escores preditos se aproximam dos escores humanos.

Outros experimentos são reportados a seguir. Na Tabela 6, treinamos o DIQA com uma base e aplicamos a versão treinada na outra base. Novamente, alguns exemplos foram selecionados apenas para observar os resultados. Nesse caso, pode-se perceber uma diferença maior entre os escores preditos e os escores esperados (humanos).

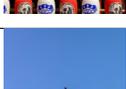
Os próximos experimentos exemplificados nas Tabelas 7, 8, 9, e 10 mostram, respectivamente, distorções por borramento, perda por codificação (JPEG), diminuição de brilho, e saturação do canal vermelho. Em todos eles. Nessas tabelas, apresentamos apenas uma imagem extraída da base KonIQ-10K e aplicamos o modelo DIQA treinado nas duas bases (LIVE IQA e KonIQ-10K), conforme indicado em cada tabela. Em todos os casos, a

Tabela 5 – Escores preditos pelo algoritmo DIQA em comparação os escores definidos por avaliadores humanos: exemplos da base KonIQ-10K.

Imagem	Escore predito	Escore humano
	0,7570	0,7901
	0,8482	0,9090
	0,7278	0,6715

Fonte: Elaborado pelo autor

Tabela 6 – Resultados da avaliação quando o DIQA é treinado em uma base de imagens e aplicado em outra.

Base de Treino	Imagem	Base da imagem	Escore predito	Escore humano
KonIQ-10K		LIVE-IQA	0,8068	0,7406
KonIQ-10K		LIVE-IQA	0,7025	0,4627
KonIQ-10K		LIVE-IQA	0,6970	0,6715
LIVE-IQA		KonIQ-10K	0,99	0,7901
LIVE-IQA		KonIQ-10K	1	0,9090
LIVE-IQA		KonIQ-10K	1	0,7257

Fonte: Elaborado pelo autor

presença de distorções gera a expectativa de queda da nota de avaliação já que refletiria uma queda de qualidade nas imagens.

Em todos os experimentos, pode-se observar uma queda no escore à medida que a imagem perde qualidade. A imagem de teste usada nessas tabelas, apresentada na Tabela 5, anteriormente, possui avaliação humana com nota 0,9090. Assim, qualquer tipo de distorção deve gerar uma nota abaixo desse valor. A queda na nota de avaliação foi maior ou menor de acordo com o nível de degradação imposta na imagem, mas o importante a ressaltar é que houve queda em todos os casos analisados, mesmo quando o DIQA foi treinado com *dataset* diferente do que a imagem analisada pertence.

Tabela 7 – Experimento com variação de borramento: foi aplicado um filtro Gaussiano de borramento, na imagem original, com diferentes níveis. A imagem da linha inferior está mais borrada que a da linha superior da tabela.

Imagem	Escore predito com DIQA treinado na LIVE IQA	Escore predito com DIQA treinado na KonIQ-10K
	0,4780	0,1892
	0,3135	0,1812

Fonte: Elaborado pelo autor

Tabela 8 – Experimento com variação perda no armazenamento em JPEG: a imagem da linha superior foi salva com 80% de perda e a da linha inferior com 90% sobre 80% de perda.

Imagem	Escore predito com DIQA treinado na LIVE IQA	Escore predito com DIQA treinado na KonIQ-10K
	0,6085	0,8420
	0,5028	0,7715

Fonte: Elaborado pelo autor

Tabela 9 – Experimento com variação de nível de brilho.

Imagem	Escore predito com DIQA treinado na LIVE IQA	Escore predito com DIQA treinado na KonIQ-10K
	0,6670	0,7673
	0,4995	0,6281

Fonte: Elaborado pelo autor

Tabela 10 – Experimento com alta saturação do canal vermelho.

Imagem	Escore predito com DIQA treinado na LIVE IQA	Escore predito com DIQA treinado na KonIQ-10K
	0,8671	0,8188

Fonte: Elaborado pelo autor

Considerando que o DIQA foi modelado para imagens de baixa resolução, antes de realizar teste com imagens com resolução diferente das imagens de treinamento, fizemos experimentos usando técnicas de janelamento e redução de resolução nas imagens de alta resolução. Dessa forma, conseguimos avaliar uma imagem de alta resolução passando como entrada para o DIQA uma imagem baixa resolução. No janelamento, a imagem de alta resolução foi dividida em várias imagens de baixa resolução. Já na redução resolução, foram consideramos 3 cenários: (1) resolução fixa em 512x384 pixels, (2) resolução em 1/2 da resolução original da imagem e (3) resolução em 1/4 da resolução original da imagem.

4.5 RESULTADOS

Esta seção apresenta os resultados dos experimentos que foram realizados nos modelos propostos no Capítulo 3, utilizando as bases de imagens LIVE, LIVE in the wild e KonIQ-10K. Os resultados são comparados com o modelos encontrados na literatura: OCPP, BRISQUE e DIQA (modelo de referência).

Nas tabelas a seguir são mostrados os resultados das medidas de cada um dos métodos

de IQA tratados nesta dissertação. A Tabela 11 mostra os resultados, utilizando a base de dados LIVE IQA. Na Tabela 12, são mostrados os resultados utilizando a base de dados LIVE in the Wild. E, por fim, na Tabela 13, são exibidos os resultados utilizando a base de dados KonIQ-10K. Os critérios utilizados para comparação e avaliação dos métodos são os mesmos descritos na Seção 4.1 Medidas para Avaliação.

Nessas tabelas, DIQA corresponde ao algoritmo original; DIQA v2 é o algoritmo sem a etapa de normalização (conforme apresentado na Seção 3.2.1) e DIQA v3 é a versão com treinamento com a base LIVE com super-resolução nas imagens da base LIVE e sem a etapa de normalização (apresentada na Seção 3.2.2). Em todas as tabelas, uma seta para cima indica que a medida é melhor o quão mais alto for seu valor e uma seta para baixo indica que valores baixos correspondem aos melhores resultados.

Tabela 11 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência utilizando a base LIVE IQA.

Algoritmo	SROCC ↑	KROCC ↑	PLCC ↑	RMSE ↓
BRISQUE	0,9368	0,7856	0,8736	0,1718
OCP	0,9925	0,9409	0,9972	0,0333
DIQA	0,9808	0,8803	0,9857	0,0487
DIQA v2	0,9480	0,8059	0,9357	0,1126
DIQA v3	0,9612	0,8319	0,9656	0,0736

Fonte: Elaborado pelo autor

Tabela 12 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência utilizando a base LIVE in the Wild.

Algoritmo	SROCC ↑	KROCC ↑	PLCC ↑	RMSE ↓
BRISQUE	0,8726	0,7608	0,8751	0,1340
OCP	0,7737	0,6885	0,8566	0,0787
DIQA	0,8678	0,6855	0,9080	0,0983
DIQA v2	0,7711	0,5822	0,8240	0,1172

Fonte: Elaborado pelo autor

Como pode ser visto nas tabelas 11, 12, 13, utilizando as imagens públicas de baixa resolução, os modelos apresentaram resultados competitivos. As medidas possuem valores muito próximos uns dos outros. A pequena diferença que existe pode ser desconsiderada dado que ainda não estamos avaliando o cenário proposto neste trabalho. Além disso, nenhum dos modelos tem uma diferença expressiva sobre os concorrentes e, por não existir unanimidade, vamos seguir comparando esses mesmos modelos, utilizando imagens de alta resolução. Também foi observado o tempo de execução para a avaliação das imagens. O OCP foi o que levou mais tempo para a avaliação das imagens. O DIQA e suas variações

Tabela 13 – SROCC, KROCC, PLCC e RMSE para o modelo DIQA de referência utilizando a base KonIQ-10K.

Algoritmo	SROCC \uparrow	KROCC \uparrow	PLCC \uparrow	RMSE \downarrow
BRISQUE	0,7625	0,6765	0,7613	0,0648
OCPD	0,8085	0,7094	0,7990	0,1694
DIQA	0,9508	0,8123	0,9631	0,0477
DIQA v2	0,7756	0,5817	0,8301	0,0830

Fonte: Elaborado pelo autor

se mostraram mais rápidos em relação aos demais modelos, como pode ser visto na Tabela 14, lembrando que não estamos considerando tempo de treinamento.

Tabela 14 – Tempo médio para avaliar uma imagem de alta resolução

Algoritmo	Tempo em segundos
BRISQUE	17,06
OCPD	428
DIQA	0,68
DIQA v2	0,71
DIQA v3	0,68

Fonte: Elaborado pelo autor

A seguir, mostramos o comportamento dos modelos quando submetidos a imagens de alta resolução. Os exemplos são de imagens capturadas de dispositivos móveis atuais e estão separados por tipo de situação. Exemplos dessas imagens podem ser vistos na Figura 8, com os escores atribuídos a cada imagem apresentados na Tabela 15.

Figura 8 – Imagens tiradas em ambiente externo com dimensões de 4640x3472 pixels.



(a)

(b)

Fonte: Compilação do autor. Acervo pessoal.

Tabela 15 – Escores das imagens da Figura 8

Algoritmo	Figura 8 (a)	Figura 8 (b)
BRISQUE	0,7821	0,7604
OCP	0,9972	0,9972
DIQA	0,9935	0,7725
DIQA v2	0,6016	0,5725
DIQA v3	0,8124	0,8921

Fonte: Elaborado pelo autor

Como pode ser visto na Tabela 15, os escores preditos pelos algoritmos para as duas imagens são muito próximos, exceto no DIQA. Neste caso, as duas imagens são de boa qualidade e em ambiente semelhantes, porém o DIQA avaliou de forma injustificável a foto 8a com escore de 0,9535, quase excelente, contudo a imagem tem qualidade mediana.

Figura 9 – Fotografias tiradas com variação de zoom com dimensões de 4640x2610 pixels. Nas imagens com zoom, é percebido uma redução na qualidade. Logo, (b) tem menor qualidade visual que (a) e (c) tem menor qualidade visual que (b)



(a) Sem zoom



(b) Com zoom



(c) Com mais zoom

Fonte: Compilação do autor. Acervo pessoal.

Tabela 16 – Escores das imagens da Figura 9

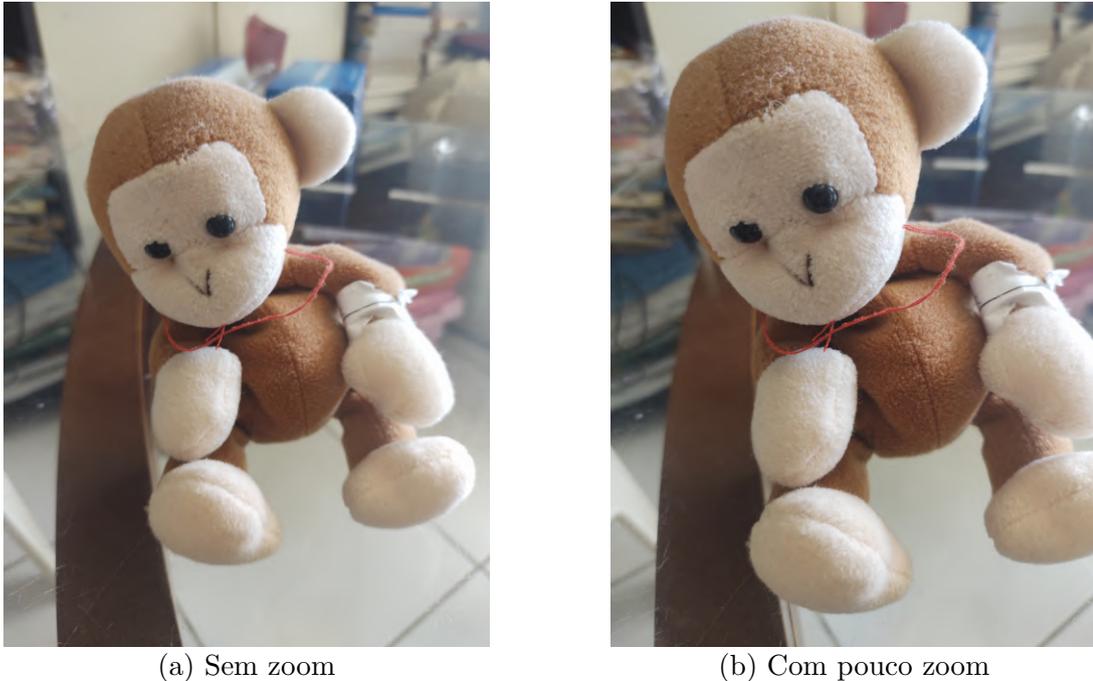
Algoritmo	Figura 9 (a)	Figura 9 (b)	Figura 9 (c)
BRISQUE	0,8824	0,5899	0,49416
OCP	0,9974	0,9962	0,9950
DIQA	0,9625	0,6872	0,4818
DIQA v2	0,6736	0,4867	0,3888
DIQA v3	0,9527	1,0000	0,6922

Fonte: Elaborado pelo autor

Como pode ser visto na Figura 9, à medida que aplicamos mais zoom, a qualidade visual da imagem é reduzida. Todos os algoritmos, exceto o DIQA v3 e o OCP, atribuíram escores coerentes. Na Tabela 16 é possível ver a redução dos escores entre as imagens das figuras Figura 9a, Figura 9b e Figura 9c. O DIQA v3, mostrou inconsistência no teste variação de zoom, atribuindo nota máxima para a imagem da Figura 9b, mesmo essa

imagem tendo pior qualidade em relação a imagem da Figura 9a. O OCPP também foi inconsistente ao apresentar uma nota tão alta e com uma variação tão pequena com a perda de qualidade.

Figura 10 – Fotografias tiradas em ambiente interno e com zoom com dimensões de 4640x3472 pixels. Em (a) o objeto em foco está mais distante logo, existe mais área borrada que em (b).



Fonte: Compilação do autor. Acervo pessoal.

Tabela 17 – Escores das imagens da Figura 10

Algoritmo	Figura 10 (a)	Figura 10 (b)
BRISQUE	0,6617	0,6742
OCPP	0,9965	0,9967
DIQA	1,000	1,000
DIQA v2	0,6769	0,6847
DIQA v3	0,7509	0,8093

Fonte: Elaborado pelo autor

Nas imagens da Figura 10, apesar do zoom aplicado na imagem 10b, essa é mais nítida e tem mais detalhes que a imagem Figura 10a. O DIQA foi o único que atribuiu escores iguais e máximos para as duas imagens, considerando-as como imagens excelentes. Os demais algoritmos atribuíram escores coerentes, com um valor maior para a imagem Figura 10b. Novamente, o OCPP mostrou uma variação muito pequena entre os escores das imagens. É importante observar que os algoritmos devem analisar as imagens como um todo. Nesses dois casos, as imagens têm um *background* com algum nível de borramento

e não deveriam apresentar escore máximo (ou perto de máximo). Nesses dois casos, o BRISQUE e as variações propostas no DIQA apresentam resultados mais próximos ao que se espera.

Figura 11 – Imagens avaliadas com o escores incoerentes pelo DIQA, todas as imagens têm dimensões de 4624x2136 pixels. Super exposição e borramento do céu em (a).



(a) escore = 1,000



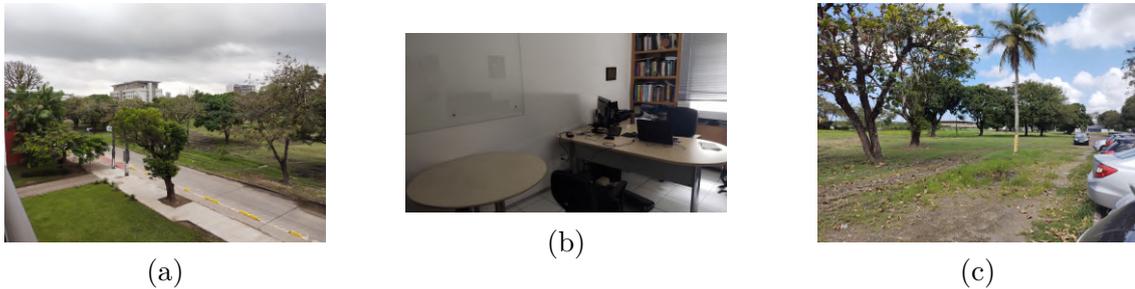
(b) escore = 0,8440

Fonte: Compilação do autor. Acervo pessoal.

Na Figura 11, o DIQA avaliou de forma contraditória as duas imagens. A imagem da Figura 11a tem problemas de super exposição de luz e nitidez, porém foi avaliada com escore máximo. Ao mesmo tempo que a imagem da Figura 11b é nítida e sem ruídos aparentes, sendo avaliada com um escore menor (0,8440).

Como pode ser visto na Figura 12, imagens com diferentes níveis de qualidade foram avaliadas com o mesmo escore pelo OCPP. A imagem da Figura 12a possui um grande nível de borramento na parte superior do céu. Na imagem Figura 12b, toda a parte de reflexo no quadro no lado esquerdo da imagem está com ruído, além da super exposição da luz na janela. Já a imagem da Figura 12c não possui nenhum tipo de ruído perceptível.

Figura 12 – Imagens avaliadas com o escore 0,9972 pelo OCPP, todas as imagens têm dimensões de 4640x3472 pixels. Alto nível de borramento na parte superior de (a). Borramento em todo o reflexo no lado esquerdo de (b).



Fonte: Compilação do autor. Acervo pessoal.

Esse fato ocorreu em 48 das 118 imagens das imagens avaliadas. Dessa forma, o OCPP não demonstrou precisão e foi incoerente na atribuição das notas dessas imagens. Esse problema foi percebido quando comparamos imagens em diferentes ambientes e ângulos.

Mesmo com ótimas medidas e avaliações nas bases de dados de imagens públicas, os modelos presentes na literatura e os modelos propostos neste trabalho mostraram inconsciências quando são avaliadas imagens de alta resolução.

Os experimentos utilizando janelamento e redução de resolução também foram incoerentes, porém não serão tratados nesse trabalho, pois é necessário maior investigação. No caso do janelamento, as inconsistências encontradas podem ser tanto no método utilizado quanto da perda de contexto pois as sub imagens são avaliadas de forma independente. Já quando aplicamos a redução de resolução, existe uma perda de dados da imagem original, que pode ter ocasionado em uma má avaliação, reduzindo a acurácia do modelo.

4.6 OUTROS EXPERIMENTOS

Nesta seção, aplicamos outros algoritmos nas imagens de alta resolução para analisar os resultados. Além dos algoritmos já apresentados nas seções anteriores, trazemos experimentos com os algoritmos NBIQA (OU; WANG; ZHU, 2019) e IL-NIQE (ZHANG; ZHANG; BOVIK, 2015).

Em (MITTAL; SOUNDARARAJAN; BOVIK, 2013), é introduzido o algoritmo *Natural Image Quality Evaluator* (NIQE). Sua ideia principal é definir medidas estatísticas que caracterizam uma imagem sem distorção. Assim, a proposta é preparar o algoritmo para lidar com imagens de alta qualidade. Com isso, ele não é submetido a nenhum tipo de distorção específica, com a intenção de poder detectar qualquer tipo de degradação. NIQE foi testado na base LIVE com bons resultados. Esses bons resultados motivaram a criação do IL-NIQE (ZHANG; ZHANG; BOVIK, 2015) (Integrated Local Natural Image Quality Evaluator). O IL-NIQE se baseia em diferentes tipos de estatísticas de cenas naturais (relacionadas à luminância, cor, contraste local, etc), alcançando resultados melhores que

o NIQE. Outro modelo baseado em estatística de cenas naturais é apresentado em (OU; WANG; ZHU, 2019): *Novel Blind Image Quality Assessment* (NBIQA). Diferente de outros modelos, o NBIQA extrai características da imagem no domínio espacial e no domínio da frequência (através da aplicação da Transformada Discreta do Cosseno na imagem de entrada). Por essa necessidade de mudança de domínio, o método é bastante custoso como é apresentado mais à frente.

Os resultados são apresentados sem normalização. Os dois algoritmos trabalham com escores de 0 a 100. Para o NBIQA, quanto maior o escore, melhor a imagem. Para o IL-NIQE, imagens de boa qualidade têm escores menores, próximos a zero.

Ambos os algoritmos foram aplicados às imagens da Figura 9, os escores atribuídos por cada um deles podem ser vistos na Tabela 18.

Tabela 18 – Escores das imagens da Figura 9 atribuídos pelo IL-NIQE e NBIQA.

Algoritmo	Figura 9 (a)	Figura 9 (b)	Figura 9 (c)
IL-NIQE	8,8550	9,722	13,441
NBIQA	55,0202	48,0704	10,8168

Fonte: Elaborado pelo autor

Como pode ser visto na Tabela 18, a degradação das imagens, à medida que o zoom é aumentado, é percebida de formas diferentes pelos algoritmos IL-NIQE e NBIQA. Ambos consideram que a imagem tem uma redução na qualidade, porém, enquanto o NBIQA considera uma queda brusca na qualidade percebida entre a imagem da Figura 9b e a da Figura 9c, no IL-NIQE é apontada uma diferença bem pequena. No geral, as avaliações do IL-NIQE geram escores com valores mais próximos a 0 que de 100, o que passa o entendimento que são imagens de boa qualidade. Nesse caso, isso não se confirma já que as três imagens possuem qualidade de mediana a ruim, principalmente a Figura 9c. Considerando o mesmo contexto, o NBIQA apesar de avaliar a Figura 9c como uma imagem ruim e com escore bem próximo a zero, não teve a mesma sensibilidade quando comparamos os escores das imagens das Figuras 9a e 9b.

Ambos os algoritmos foram executados com código implementado em Matlab, disponibilizado pelos autores. Com essa implementação, o tempo médio de execução do NBIQA (que depende de conversão para outro domínio através ds DCT) foi de 8 minutos por imagem. Já o IL-NIQE, precisou de 6 segundos por imagem, em média.

5 CONCLUSÃO

Este trabalho analisou métodos de avaliação de qualidade de imagem baseados em modelos de *Deep Learning*. O objetivo do trabalho foi avaliar algoritmos de avaliação de qualidade de imagem encontrados na literatura, como o DIQA, OCPP e BRISQUE. Além disso, foram propostas duas variações do modelo DIQA, o DIQA v2 e DIQA v3. Esses modelos apresentam mudanças que poderiam gerar melhores resultados ao modelo original quando aplicados em imagens de alta resolução. Para todos os modelos que necessitaram de treinamento, foram utilizadas bases de dados de imagens públicas. As bases utilizadas foram: LIVE IQA, LIVE in the Wild e KonIQ-10K.

Sendo assim, ao longo deste estudo, foram apresentados diversos experimentos que visam avaliar os modelos existentes de avaliação de qualidade de imagem sem referência no contexto de imagens de alta resolução capturadas de dispositivos móveis. Esse contexto foi escolhido visto que não existe nenhuma base de dados pública com imagens de alta resolução rotuladas em termos de qualidade de imagem.

Os experimentos mostraram que, apesar de possuírem ótimos resultados na literatura, os algoritmos BRISQUE, OCPP e DIQA apresentam inconsistências quando são aplicados imagens de alta resolução.

Além disso, o OCPP e o BRISQUE possuem um tempo de avaliação de imagem muito alto, algo que pode limitar o uso desses métodos em aplicações práticas.

O OCPP mostrou pouca variação no score quando avaliamos imagens de alta resolução, algo que pode ser um problema para eleger que uma imagem seja melhor que outra, visto a variação tão pequena as imagens podem ser consideradas similares mesmo com qualidades muito distintas. Nos casos avaliados, a diferença do score ocorreu a partir da terceira casa decimal.

Nos experimentos, o DIQA não mostrou sensibilidade quando as imagens tinham boa qualidade, em muitos casos foram atribuídos scores máximos, classificando as imagens como excelentes e iguais mesmo em imagens com nível de qualidade diferente.

Na maioria dos experimentos feitos, os algoritmos avaliados atribuíram scores coerentes quando comparados com scores de outras imagens no mesmo ângulo e cenário. Porém, quando foram avaliadas imagens em cenários distintos, todos os algoritmos apresentaram algum tipo de inconsistência, como visto no capítulo anterior.

5.1 TRABALHOS FUTUROS

Para dar continuidade ao trabalho de pesquisa descrito nesta dissertação, sugere-se, nesta seção, propostas de trabalhos futuros a serem realizadas como, por exemplo, modificações na arquitetura convolucional do DIQA de modo a extrair mais características capazes de

corrigir os problemas encontrados e o escore mais sensível para imagens de alta resolução. Além disso, sugere-se criar uma base de dados pública, com imagens de alta resolução capturadas de dispositivos móveis e rotuladas com escore de qualidade para poder ampliar os recursos disponíveis para de treinamento dos modelos de *Deep learning*. O uso *crowdsourcing* pode facilitar essa criação.

REFERÊNCIAS

- ALTMAN, N. S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, Taylor & Francis, v. 46, n. 3, p. 175–185, 1992.
- BRAHNAM, S.; JAIN, L. C.; LUMINI, A.; NANNI, L. Introduction to local binary patterns: new variants and applications. In: *Local Binary Patterns: New Variants and Applications*. [S.l.]: Springer, 2014. p. 1–13.
- BREIMAN, L. Random forests. *Machine learning*, Springer, v. 45, n. 1, p. 5–32, 2001.
- BULL, D. *Chapter 10-measuring and managing picture quality*. [S.l.]: Academic Press Oxford, 2014. 317–360 p.
- BURGES, C. J. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, Springer, v. 2, n. 2, p. 121–167, 1998.
- CHAI, T.; DRAXLER, R. R. Root mean square error (rmse) or mean absolute error (mae). *Geoscientific Model Development Discussions*, v. 7, n. 1, p. 1525–1534, 2014.
- CHANG, C.-C.; LIN, C.-J. Libsvm: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, Acm New York, NY, USA, v. 2, n. 3, p. 1–27, 2011.
- DOZAT, T. Incorporating nesterov momentum into adam. 2016.
- FREITAS, P. G.; AKAMINE, W. Y.; FARIAS, M. C. No-reference image quality assessment using orthogonal color planes patterns. *IEEE Transactions on Multimedia*, IEEE, v. 20, n. 12, p. 3353–3360, 2018.
- GHADIYARAM, D.; BOVIK, A. Live in the wild image quality challenge database. *Online: <http://live.ece.utexas.edu/research/ChallengeDB/index.html>*[Mar, 2017], 2015.
- GHADIYARAM, D.; BOVIK, A. C. Massive online crowdsourced study of subjective and objective picture quality. *IEEE Transactions on Image Processing*, IEEE, v. 25, n. 1, p. 372–387, 2015.
- GOLESTANEH, S. A.; CHANDLER, D. M. No-reference quality assessment of jpeg images via a quality relevance map. *IEEE signal processing letters*, IEEE, v. 21, n. 2, p. 155–158, 2013.
- GU, K.; LIN, W.; ZHAI, G.; YANG, X.; ZHANG, W.; CHEN, C. W. No-reference quality metric of contrast-distorted images based on information maximization. *IEEE transactions on cybernetics*, IEEE, v. 47, n. 12, p. 4559–4565, 2016.
- GU, K.; WANG, S.; YANG, H.; LIN, W.; ZHAI, G.; YANG, X.; ZHANG, W. Saliency-guided quality assessment of screen content images. *IEEE Transactions on Multimedia*, IEEE, v. 18, n. 6, p. 1098–1110, 2016.
- GU, K.; ZHAI, G.; YANG, X.; ZHANG, W.; CHEN, C. W. Automatic contrast enhancement technology with saliency preservation. *IEEE Transactions on Circuits and Systems for Video Technology*, IEEE, v. 25, n. 9, p. 1480–1494, 2014.

- He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. p. 770–778. ISSN 1063-6919.
- 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.
- KENDALL, M. G. The treatment of ties in ranking problems. *Biometrika*, JSTOR, v. 33, n. 3, p. 239–251, 1945.
- KIM, J.; LEE, S. Fully deep blind image quality predictor. *IEEE Journal of selected topics in signal processing*, IEEE, v. 11, n. 1, p. 206–220, 2016.
- KIM, J.; NGUYEN, A.-D.; LEE, S. Deep cnn-based blind image quality predictor. *IEEE transactions on neural networks and learning systems*, IEEE, v. 30, n. 1, p. 11–24, 2018.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- LASMAR, N.-E.; STITOU, Y.; BERTHOUMIEU, Y. Multiscale skewed heavy tailed model for texture analysis. In: IEEE. *2009 16th IEEE International Conference on Image Processing (ICIP)*. [S.l.], 2009. p. 2281–2284.
- LECUN, Y.; BENGIO, Y. The handbook of brain theory and neural networks. In: ARBIB, M. A. (Ed.). Cambridge, MA, USA: MIT Press, 1998. cap. Convolutional Networks for Images, Speech, and Time Series, p. 255–258. ISBN 0-262-51102-9. Disponível em: <http://dl.acm.org/citation.cfm?id=303568.303704>.
- LI, L.; ZHOU, Y.; LIN, W.; WU, J.; ZHANG, X.; CHEN, B. No-reference quality assessment of deblocked images. *Neurocomputing*, Elsevier, v. 177, p. 572–584, 2016.
- LI, Q.; LIN, W.; FANG, Y. Bsd: Blind image quality assessment based on structural degradation. *Neurocomputing*, Elsevier, v. 236, p. 93–103, 2017.
- LI, Q.; LIN, W.; XU, J.; FANG, Y. Blind image quality assessment using statistical structural and luminance features. *IEEE Transactions on Multimedia*, IEEE, v. 18, n. 12, p. 2457–2469, 2016.
- LIM, B.; SON, S.; KIM, H.; NAH, S.; LEE, K. M. Enhanced deep residual networks for single image super-resolution. In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. [S.l.: s.n.], 2017. p. 136–144.
- LIU, L.; DONG, H.; HUANG, H.; BOVIK, A. C. No-reference image quality assessment in curvelet domain. *Signal Processing: Image Communication*, Elsevier, v. 29, n. 4, p. 494–505, 2014.
- MITTAL, A.; MOORTHY, A. K.; BOVIK, A. C. Blind/referenceless image spatial quality evaluator. In: IEEE. *2011 conference record of the forty fifth asilomar conference on signals, systems and computers (ASILOMAR)*. [S.l.], 2011. p. 723–727.
- MITTAL, A.; MOORTHY, A. K.; BOVIK, A. C. No-reference image quality assessment in the spatial domain. *IEEE Transactions on image processing*, IEEE, v. 21, n. 12, p. 4695–4708, 2012.

- MITTAL, A.; MOORTHY, A. K.; BOVIK, A. C.; CHEN, C. W.; CHATZIMISIOS, P.; DAGIUKLAS, T.; ATZORI, L. No-reference approaches to image and video quality assessment. *Multimedia Quality of Experience (QoE): Current Status and Future Requirements*, John Wiley & Sons, v. 99, 2015.
- 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.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Icml*. [S.l.: s.n.], 2010.
- OJALA, T.; PIETIKÄINEN, M.; MÄENPÄÄ, T. Gray scale and rotation invariant texture classification with local binary patterns. In: SPRINGER. *European conference on computer vision*. [S.l.], 2000. p. 404–420.
- OU, F.-Z.; WANG, Y.-G.; ZHU, G. A novel blind image quality assessment method based on refined natural scene statistics. In: *2019 IEEE International Conference on Image Processing (ICIP)*. [S.l.: s.n.], 2019. p. 1004–1008.
- QUOTEV. *Imagem escura na floresta*. 2022. Disponível em: <https://i.quotev.com/muffyvmig2ca.jpg>. Acesso em 20 jun. 2022.
- RUDERMAN, D. L. The statistics of natural images. *Network: computation in neural systems*, IOP Publishing, v. 5, n. 4, p. 517, 1994.
- 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.
- SCHÖLKOPF, B.; SMOLA, A. J.; WILLIAMSON, R. C.; BARTLETT, P. L. New support vector algorithms. *Neural computation*, MIT Press One Rogers Street, Cambridge, MA 02142-1209, USA journals-info . . . , v. 12, n. 5, p. 1207–1245, 2000.
- SHARIFI, K.; LEON-GARCIA, A. Estimation of shape parameter for generalized gaussian distributions in subband decompositions of video. *IEEE Transactions on Circuits and Systems for Video Technology*, IEEE, v. 5, n. 1, p. 52–56, 1995.
- SHEIKH, H. R.; BOVIK, A. C. Image information and visual quality. *IEEE Transactions on image processing*, IEEE, v. 15, n. 2, p. 430–444, 2006.
- SHEIKH, H. R.; SABIR, M. F.; BOVIK, A. C. A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on image processing*, IEEE, v. 15, n. 11, p. 3440–3451, 2006.
- SHEIKH, H. R.; WANG, Z.; CORMACK, L.; BOVIK, A. C. Live image quality assessment database release 2 (2005). URL <http://live.ece.utexas.edu/research/quality>, 2005.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- TANG, H.; JOSHI, N.; KAPOOR, A. Learning a blind measure of perceptual image quality. In: IEEE. *CVPR 2011*. [S.l.], 2011. p. 305–312.

- YE, P.; KUMAR, J.; KANG, L.; DOERMANN, D. Unsupervised feature learning framework for no-reference image quality assessment. In: IEEE. *2012 IEEE conference on computer vision and pattern recognition*. [S.l.], 2012. p. 1098–1105.
- ZHANG, L.; ZHANG, L.; BOVIK, A. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing*, IEEE, v. 24, n. 12, p. 2579–2591, 2015.
- ZHANG, M.; MURAMATSU, C.; ZHOU, X.; HARA, T.; FUJITA, H. Blind image quality assessment using the joint statistics of generalized local binary pattern. *IEEE Signal Processing Letters*, IEEE, v. 22, n. 2, p. 207–210, 2014.