Pós-Graduação em Ciência da Computação

Daniela de Sousa Costa

Aprimoramento de imagens baseado em estimativa de iluminante e técnicas de Aprendizagem Profunda

Universidade Federal de Pernambuco
posgraduacao@cin.ufpe.br
http://cin.ufpe.br/~posgraduacao

Recife
2020

Daniela de Sousa Costa

Aprimoramento de imagens baseado em estimativa de iluminante e técnicas de Aprendizagem Profunda

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: inteligência computacional
Orientador: Prof. Dr. Carlos Alexandre Barros de Mello

Recife
2020

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

C837a Costa, Daniela de Sousa
Aprimoramento de imagens baseado em estimativa de iluminante e técnicas de aprendizagem profunda / Daniela de Sousa Costa. – 2020.
75 f.: il., fig., tab.

Orientador: Carlos Alexandre Barros de Mello.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2020.
Inclui referências e apêndice.

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

006.31 CDD (23. ed.) UFPE - CCEN 2020 -89

Daniela de Sousa Costa

“Aprimoramento de imagens baseado em estimativa de iluminante e técnicas de Aprendizagem Profunda”

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: 27 de fevereiro de 2020.

BANCA EXAMINADORA

_____

Prof. Dr. Adriano Lorena Inácio de Oliveira
Centro de Informática / UFPE

_____

Prof. Dr. Wellington Pinheiro dos Santos
Departamento de Engenharia Biomédica/ UFPE

_____

Prof. Dr. Carlos Alexandre Barros de Mello
Instituto Federal de Pernambuco

Dedico este trabalho Àquele que me dá a razão de viver. Amo a sua Palavra, pois é lâmpada para os meus pés e luz para os meus caminhos.

AGRADECIMENTOS

Aos meus pais maravilhosos, por todo amor, confiança e apoio oferecidos a mim incondicionalmente. Sou grata a Deus por ser filha de vocês.

Aos meus queridos irmãos Fredson, Adriano e Sueline, por todos os momentos felizes que vivemos e pelos que ainda virão.

À Primeira Igreja Batista da Cidade Universitária, por me acolher tão bem e ser a minha família em Recife.

Ao meu querido professor Carlos Alexandre, por todo ensino e paciência a mim dedicados durante esses anos. Sei que ainda tenho muito a aprender com ele.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo suporte financeiro dado através da bolsa de Mestrado.

RESUMO

Uma cena quando capturada por dispositivos pode apresentar diferenças significativas entre aquilo que é observado diretamente pelo olho humano e sua representação na forma de imagem. Isto se deve à capacidade que os seres humanos têm de perceber certos aspectos da imagem, como cor e detalhes em regiões escuras independentemente da iluminação. A implementação de tais habilidades em sistemas computacionais se mostra benéfica em várias aplicações gráficas e de visão computacional, tais como as que envolvem classificação, segmentação semântica e renderização de cenas. Neste trabalho, são abordados dois tipos de aprimoramento de imagem. O primeiro visa corrigir as cores dos objetos de uma cena, de maneira que as mesmas possam ser identificadas corretamente independentemente da cor do iluminante utilizado para captura, propriedade conhecida como constância de cor. Já o segundo tipo de aprimoramento é voltado para casos onde a captura da imagem é feita sob condições de baixa luminosidade. Para ambos os problemas, percebeu-se que o ponto central é a influência da iluminação que pode gerar efeitos não desejáveis sobre a cena. A partir dessa observação, são apresentados dois métodos baseados em redes neurais convolucionais que, ao receberem uma imagem, estimam o iluminante sendo este utilizado para correção da mesma. Experimentos revelam que as estratégias propostas são capazes de proporcionar resultados compatíveis e, em certos casos, superiores aos algoritmos do estado da arte.

Palavras-chaves: Constância de cor. Aprimoramento de imagens. Aprendizagem profunda.

ABSTRACT

A scene captured by devices can present significant differences between objects that are directly visible to the human eye and their representation as an image. This ability allows humans to perceive certain aspects of the image, such as color and details in dark regions, somewhat independently of lighting. The implementation of such skills in computer systems shows benefits in various graphics and computer vision applications, such as classification tasks, semantic segmentation, and scene rendering. In this work, two types of image enhancement are introduced. The first method for image enhancement aims to correct the colors of the objects in a scene so that they can be correctly identified regardless of the color of the light source used to capture the image, a property known as color constancy. The second enhancement method focuses on cases where the image is captured under low-light conditions. For both problems, the central point is the influence of lighting that can generate undesirable effects on the scene. For this reason, two methods are presented based on convolutional neural networks that receive an image and estimate the illuminant that is used to correct the scene. Experimental results reveal that the proposed methods achieve compatible results and, in some cases, demonstrate superior performance to the state-of-the-art methods.

Keywords: Color constancy. Image enhancement. Deep learning.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

SR	*Spectral Residual*
DSR	*Dense and Sparse Reconstruction*
GR	*Graph-regularized Saliency Detection with Convex-hull-based Center Prior*
MC	*Saliency Detection via Absorbing Markov Chain*
CNN	*Convolutional Neural Network*
ReLU	*Rectified Linear Units*
FC4	*Fully Convolutional Color Constancy with Confidence-weighted Pooling*
DSLR	*Digital Single-Lens Reflex*
MCC	*Macbeth ColorChecker*
HE	*Histogram Equalization*
CLAHE	*Contrast Limited Adaptive Histogram Equalization*
LDR	*Layered Difference Representation*
SRIE	*Simultaneous Reflectance and Illumination Estimation*
LIME	*Low-light Image Enhancement*
BM3D	*Block-matching and 3D filtering*
JPEG	*Joint Photographics Experts Group*
DnCNN	*Denoising Convolutional Neural Network*
WIIEN	*Weakly Illuminated Image Enhancement Network*
Conv	Camada de Convolução
BN	*Batch Normalization*
SSIM	*Structural Similarity Index*
PSNR	*Peak Signal-to-Noise Ratio*
MSE	*Mean Squared Error*
LOE	*Lightness Order Error*

SUMÁRIO

1 INTRODUÇÃO

Neste tempo em que há um crescente interesse na captura e compartilhamento de fotos em redes sociais, a busca pelo aumento da qualidade das imagens tem motivado não somente a indústria responsável pela entrega de dispositivos com sensores mais eficientes, como também impulsiona o campo de processamento de imagens voltado para o aprimoramento de imagens. Isto porque as fotos capturadas por câmeras profissionais ou por celular podem sofrer diversos tipos de influências que dificultam ou mesmo inviabilizam a visualização de detalhes da cena. As condições que frequentemente degradam a imagem incluem más condições de tempo, pouca ou inadequada iluminação, objetos em movimento, entre outras. Tudo isso compromete diretamente a qualidade da imagem (RAHMAN et al., 2018).

Entende-se por aprimoramento de imagem técnicas capazes de destacar ou corrigir características de uma cena consideradas importantes para uma aplicação específica. Por exemplo, o aperfeiçoamento do contraste pode ser um pré-processamento importante para aplicações como detecção de borda ou segmentação. A Figura 1 ilustra esse último caso, onde sobre uma imagem com boa iluminação e sua versão escurecida (com pouco contraste) é aplicado o algoritmo de segmentação Mean Shift (CHENG, 1995) (com os mesmos parâmetros para ambas as imagens). Os resultados (Figuras 1c e 1d) indicam que a segmentação com Mean Shift é menos precisa quando aplicada em condições de baixa iluminação.

Figura 1 – Aplicação de método Mean Shift em diferentes condições de iluminação.

(a) Imagem original (b) Versão escurecida (c) Segmentação em (a) (d) Segmentação em (b)

Fonte: O autor.

Embora existam vários aspectos onde o aprimoramento de imagens possa focar, geralmente esses métodos se concentram em atributos específicos, como aprimoramento de contraste ou constância de cor (PITKÄNEN, 2019). Normalmente, o objetivo desses métodos é exibir as informações da imagem de forma mais clara e fiel possível.

O uso de técnicas digitais para o aprimoramento da qualidade de imagens tem crescido significativamente nos últimos anos. É possível perceber esse crescimento no uso de métodos de processamento de imagens coloridas em áreas de pesquisas médicas, processamento de imagens aéreas e nas indústrias de impressão (KOSCHAN; ABIDI, 2008).

Imagens capturadas por câmeras digitais teoricamente são hábeis para descrever informações similares às cenas reais. No entanto, alguns fatores podem torná-las diferentes, como as configurações do dispositivo, limitação do sensor, ângulo da captura, diferentes condições de iluminação, entre outros. A Figura 2 ilustra um caso típico onde a mesma cena foi capturada com ângulo e luz diferentes. Percebe-se que devido a esses fatores (e provavelmente outros, como a qualidade do sensor) as cores dos prédios apresentam uma certa variabilidade. Para resolver a questão da diferença de cores, é necessário aplicar um algoritmo para correção de cores. Essa correção torna as imagens aprimoradas mais adequadas para uso em sistemas de visão computacional (TIAN, 2018).

Figura 2 – Exemplo de uma mesma cena capturada em diferentes ângulos e condições de iluminação.

Fonte: Praça do Marco Zero, Recife.

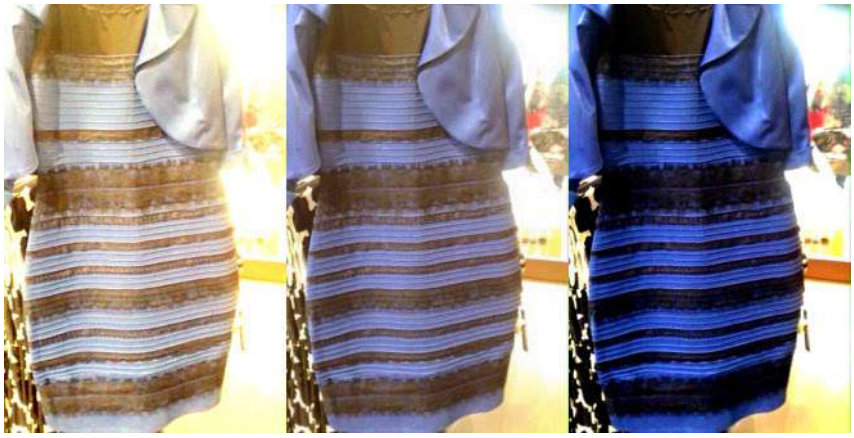
Abordagens de constância de cor buscam a eliminação dos efeitos da fonte de luz sobre as cores dos objetos. Geralmente, esse processo consiste em duas etapas. A primeira envolve a estimativa dos parâmetros do iluminante da cena, enquanto que, na segunda etapa, as cores corrigidas dos objetos são calculadas sob a iluminação estimada. O objetivo, então, é corrigir a imagem de forma a parecer que a mesma foi capturada sob luz branca, a qual é geralmente adotada como iluminação canônica (FORSYTH, 1990). Isso porque,

essa é a iluminação para a qual muitas vezes a câmera está calibrada (BARNARD et al., 2002).

As cores presentes em imagens são determinadas por propriedades intrínsecas das superfícies dos objetos, bem como pela cor da fonte de luz. Sistemas computacionais robustos que consideram as cores devem filtrar quaisquer efeitos do iluminante sobre a cena, ou seja, é desejável que tenham habilidade para proporcionar constância de cor sobre as imagens de entrada. O sistema visual humano tem naturalmente essa habilidade de corrigir os efeitos da cor da fonte de luz, embora que ainda não se saiba com detalhes como funciona esse mecanismo no corpo humano (GIJSENIJ; GEVERS; WEIJER, 2011).

No entanto, essa forma de percepção das cores independentemente das condições de iluminação pode falhar mesmo para seres humanos. Um exemplo disso é o caso do vestido (ver Fig. 3), que gerou polêmica em 2015 por dividir a opinião pública quanto às suas cores. Parte das pessoas acreditava ver azul e preto e outra parte, branco e dourado. Diversas explicações para essa divergência foram levantadas, mas o que ficou claro é que, de fato, a luminosidade pode interferir diretamente na percepção de cores.

Figura 3 – Vestido que dividiu opiniões na internet. A versão original está no centro da imagem, já as demais foram adaptadas a fim de ilustrar as diferentes percepções de cores possíveis.



Fonte: (GOMES, 2017).

A cor é uma característica importante para descrição de informações de objetos em cenas. Não obstante, o contraste das imagens também é de significativa relevância para expressar informações estruturais da cena. Novamente, a forma como essas cenas são capturadas pode fazer com que as imagens sejam apresentadas com níveis de contraste baixos. Fatores como iluminação ruim, câmeras de baixa qualidade ou com configurações inadequadas para a fonte de luz usada podem diminuir a qualidade da imagem capturada. Tais condições muitas vezes impedem que alguns conteúdos importantes de uma cena estejam claramente visíveis (TIAN, 2018).

Métodos de aprimoramento de imagem baseados em melhoria de contraste são realmente úteis para aplicações que exigem imagens de entrada com detalhes de textura

diferenciados e cores otimizadas. Sistemas que trabalham com imagens médicas, rastreamento de objetos e detecção de rosto são exemplos de aplicações onde a extração de conteúdo significativo pode exigir uma etapa de realce das características da cena (RAHMAN et al., 2018). Um exemplo desse tipo de aprimoramento é ilustrado na Figura 4, onde o rosto humano, por estar em uma região escura, não tem seus detalhes característicos (olhos, nariz, etc.) bem visíveis.

Figura 4 – Exemplo de aprimoramento de imagem por contraste.

Fonte: Adaptado de Tian (2018).

Diferentes tipos de metodologias têm sido propostas para o aprimoramento do contraste de uma imagem. Uma das mais populares é a baseada em histogramas. Os métodos dessa categoria são divididos em globais e locais; geralmente, exigem uma seleção cuidadosa de parâmetros para não resultar em problemas de aprimoramento excessivo. Normalmente, métodos de aprimoramento global produzem efeitos de superaprimoramento nas regiões brilhantes da cena. Já os de aprimoramento local comumente obtêm resultados de subaprimoramento nas regiões mais escuras (TIAN, 2018). Na figura 5 é possível ver alguns desses efeitos destacados pelo retângulo vermelho.

Figura 5 – Exemplo de aprimoramento de contraste por método local e global.

(a) Imagem de baixo contraste (b) Resultado de método global (c) Resultado de método local

Fonte: Adaptado de Tian (2018).

A inteligência artificial, mais especificamente as técnicas de aprendizagem profunda, tem sido aplicada com sucesso em problemas referentes a melhoria da qualidade de ima-

gens (PITKÄNEN, 2019). Trabalhos anteriores sobre aprimoramento automático de imagens já vêm incluindo aprendizagem profunda, como em (CHEN; XU; KOLTUN, 2017) onde redes totalmente convolucionais são treinadas a fim de melhorar a qualidade de fotos capturadas por dispositivos móveis.

1.1 OBJETIVOS

O objetivo geral desta pesquisa consiste em propor duas técnicas de aprimoramento de imagens baseadas em estimativas de iluminante e aprendizagem profunda. Sendo a primeira voltada para constância de cor e a segunda, para o aprimoramento de imagens de baixa luminosidade.

1.2 OBJETIVOS ESPECÍFICOS

Durante o desenvolvimento desta pesquisa os seguintes objetivos específicos são alcançados:

1. Investigação dos métodos de estimativa de iluminante clássicos e modernos;
2. desenvolvimento de melhorias no desempenho de métodos voltados para o problema de constância de cor;
3. desenvolvimento de metodologia para aprimoramento de imagens fracamente iluminadas;
4. validação dos métodos propostos através da aplicação das medidas mais comumente encontradas em trabalhos do estado da arte.

1.3 ORGANIZAÇÃO DA DISSERTAÇÃO

Esta dissertação é constituída por seis capítulos. O segundo capítulo aborda alguns dos conceitos básicos envolvidos nesta pesquisa como, por exemplo, a percepção e constância de cor. O terceiro capítulo apresenta métodos do estado da arte para constância de cor e para aprimoramento de imagens capturadas sob baixa iluminação, bem como técnicas mais recentes que utilizam aprendizagem profunda. O quarto capítulo, por sua vez, propõe duas metodologias para o aprimoramento de imagens. A primeira consiste em duas modificações feitas sobre uma abordagem já existente para constância de cor, enquanto que a segunda é voltada para o aprimoramento de imagens fracamente iluminadas. O quinto capítulo, de experimentos, analisa e compara o desempenho de cada metodologia proposta neste trabalho com os resultados obtidos pelos métodos do estado da arte. Por fim, o sexto capítulo expõe as devidas conclusões, contribuições e as perspectivas de trabalhos futuros.

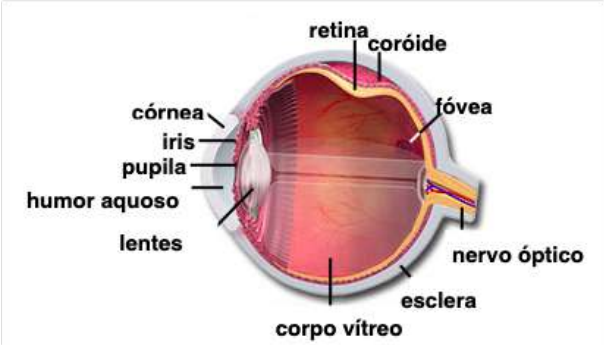
2 CONCEITOS BÁSICOS

Neste capítulo são abordados alguns dos conceitos considerados importantes para uma melhor compreensão deste trabalho, sendo eles: anatomia do olho humano, percepção de cores, constância de cor, mapas de saliência e aprendizagem profunda.

2.1 ANATOMIA DO OLHO HUMANO

O olho humano é o órgão sensorial da visão e é por ele que começa o processamento de informação do sistema visual humano (EBNER, 2007). Sua estrutura, ilustrada na Figura 6, consiste de vários componentes ópticos, tais como córnea, íris, pupila, corpo vítreo, nervo óptico e retina. Quando um objeto é observado, primeiramente ele é focalizado através da córnea e da lente. O processo de convergência dos raios de luz refletidos do objeto em direção ao olho começa na córnea, tecido transparente e que não possui vasos sanguíneos. Mais internamente ao globo ocular, encontra-se a íris, um diafragma circular responsável por controlar a quantidade de luz que entra no olho, protegendo assim as demais estruturas oculares de níveis prejudiciais de iluminação.

Figura 6 – Principais estruturas do olho humano.



Fonte: Adaptado do site Olympus Corporation [1].

Atrás da íris, está a lente natural do olho (também chamada de cristalino). Essa estrutura incolor, que mede cerca de 4 mm de espessura e 9 mm de diâmetro, auxilia no poder de foco da visão. Já o humor vítreo ou corpo vítreo, localizado após o cristalino, consiste em uma substância gelatinosa e viscosa que ocupa a maior parte do espaço do globo ocular, sendo usado para manter a pressão do olho constante (NETTO, 2003).

Os raios de luz, ao entrarem no sistema óptico, precisam passar por essas e outras estruturas internas do olho. Ao percorrerem esse caminho, sofrem um desvio produzindo uma imagem invertida do objeto na superfície da retina. A partir dessa camada, a informação sob aspecto de impulsos elétricos é enviada ao cérebro, onde são feitas as correções de

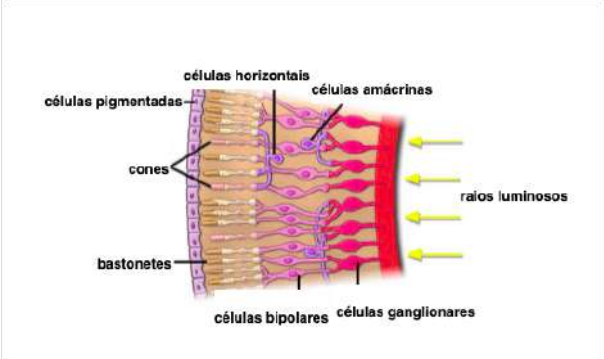
[1] Disponível em: <https://www.olympus-lifescience.com/en/microscope-resource/primer/lightandcolor/humanvisionintro/>. Acessado em 22 de Janeiro de 2020.

orientação. Uma vez que é grande a importância da retina para o sistema visual humano, a próxima seção é dedicada ao detalhamento desse componente.

2.1.1 Retina

A retina humana é um tecido fino sensitivo (com média de 0,5 μm de espessura) composto por milhares de células fotossensíveis e que reveste ceca de 70% da superfície interna ocular (GRAND, 1957). Basicamente, consiste em um conjunto de receptores que mensuram a quantidade de luz incidente, podendo ser dividida em três principais camadas, sendo cada uma constituída por um tipo de célula. Adotando o sentido da parte mais externa da retina para a mais interna, são encontradas as células ganglionares sendo seguidas pela camada de células bipolares e, por fim, pela camada das células fotossensíveis (como ilustra a Figura 7).

Figura 7 – Anatomia da retina.

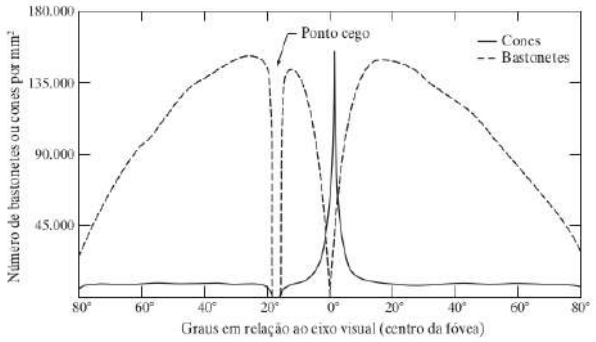


Fonte: Adaptado do site Olympus Corporation [1].

Existem dois tipos de células fotossensíveis, chamadas de cones e bastonetes. De forma simplificada, os cones são capazes de fazer distinções entre as cores, sendo usados em condições de alta luminosidade (visão fotópica). Já os bastonetes, por serem capazes de fazer diferenciação das tonalidades de claro e escuro em ambientes pouco iluminados, são de extrema importância para visão escotópica (EBNER, 2007).

Os bastonetes são mais numerosos que os cones e estão espalhados por boa parte da retina, com exceção da região da fóvea onde estão concentrados os cones (ver Figura 8). Isto faz com que esta região apresente maior acuidade visual, ou seja, nela há uma maior capacidade do olho para distinção de detalhes como forma e contorno de objetos.

Figura 8 – Distribuição de cones e bastonetes na retina.

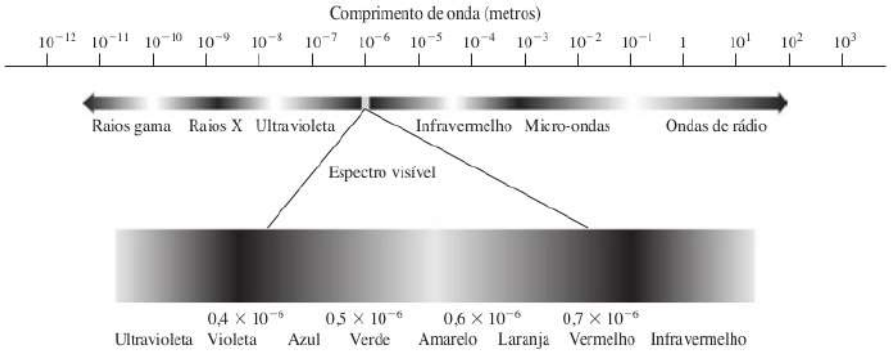


Fonte: (GONZALEZ; WOODS, 2006)

2.2 PERCEPÇÃO DE CORES

A percepção de cores ocorre como resposta do sistema nervoso quando estimulado por diferentes comprimentos de onda de luz que incidem sobre a retina. A luz é uma forma de radiação eletromagnética cuja frequência é visível ao olho humano (GONZALEZ; WOODS, 2006). O espectro visível (de cores) corresponde a uma pequena fração do espectro eletromagnético, o qual inclui ondas de rádio, micro-ondas, raios infravermelhos e raios X. A Figura 9 mostra que a banda visível do espectro eletromagnético começa aproximadamente de $0,43\mu m$ (cor violeta) e vai até $0,79\mu m$ (vermelho).

Figura 9 – Espectro eletromagnético com destaque para a faixa que é visível ao ser humano.



Fonte: (GONZALEZ; WOODS, 2006)

A cor percebida pelo homem em um objeto diz respeito às propriedades físico-químicas da superfície deste em refletir determinados comprimentos de onda. Por exemplo, uma superfície só apresentará cor azul se refletir ondas de luz com comprimentos na faixa entre 450 e 495 μm.

Toda cor possui três características principais responsáveis geralmente pela sua distinção das demais cores. A primeira propriedade é chamada de matiz que se refere à cor dominante percebida por um observador. Já a segunda é a saturação, que mede o grau

de pureza ou à quantidade de luz branca misturada ao matiz. As cores puras do espectro, por exemplo, são totalmente saturadas. Enquanto isso, o brilho refere-se a capacidade da cor em refletir a luz branca, tornando o matiz mais claro ou mais escuro.

A explicação para a visão das cores pode ser respondida por duas teorias principais: a teoria tricromática (também conhecida como teoria de Young-Helmholtz) e a de processos oponentes. Inicialmente, vistas como excludentes, atualmente são consideradas complementares, pois explicam processos que acontecem em diferentes níveis do sistema visual humano.

2.2.1 Teoria tricromática

Thomas Young and Hermann von Helmholtz propuseram a teoria tricromática a qual sustenta a hipótese de que a percepção de cores é baseada em três diferentes de cones que são especialmente sensíveis às ondas longas, médias e curtas de luz. Cada onda corresponde a percepção das cores vermelho, verde e azul, respectivamente. A junção dessas três cores primárias poderia representar todas as demais do espectro visível (KOSCHAN; ABIDI, 2008).

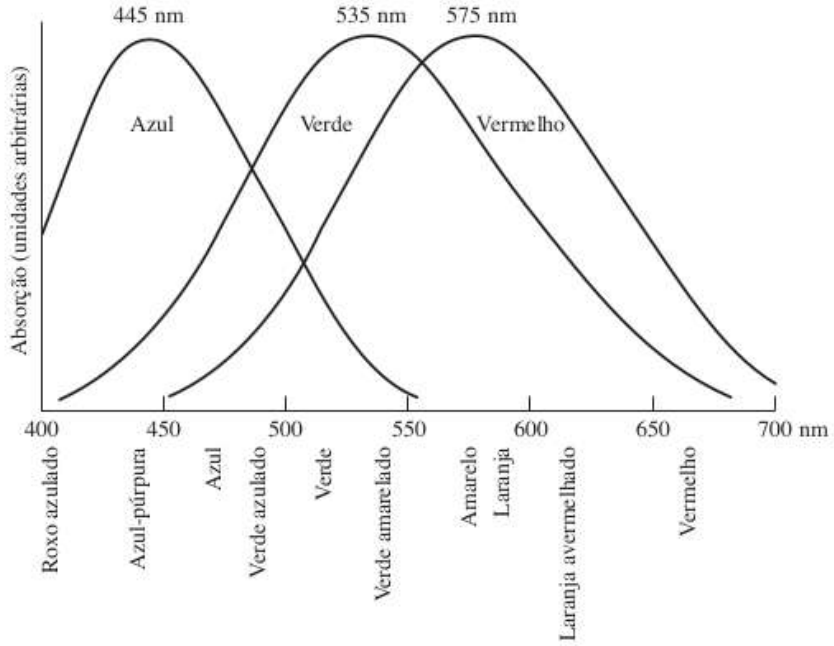
Essa teoria começou quando Young propôs que aquilo que o ser humano percebia como cor era resultado da ativação de três diferentes tipos de receptores que o olho teria. Mais tarde, Helmholtz prosseguiu com experimentos sobre essa teoria, descobrindo que pessoas com visão de cores normal podem perceber cores diversas através da combinação de apenas três comprimentos de onda de luz. Cada receptor está ligado a um conjunto próprio de nervos que enviam a mensagem ao sistema visual. Os três tipos de mensagens são, então, combinadas e produzindo a percepção de uma única cor.

Posteriormente, experimentos confirmaram que os cones presentes na retina podem ser divididos em três grupos principais, cada um capaz de detectar cores correspondentes ao vermelho, ao verde e ao azul. Na Figura 10 são mostradas as curvas de absorção da luz para tipo de cone. O máximo de absorvimento por parte dos que são mais sensíveis aos espectros vermelhos, verdes e azuis é de, aproximadamente, 575 nm, 535nm e 445nm, respectivamente.

A aplicação do princípio dessa teoria é comumente encontrada em televisões, celulares, câmeras fotográficas, monitores de computador e outros equipamentos eletrônicos nos quais a representação de cada pixel colorido requer três subpixels (vermelho, verde e azul). De forma que se estes estão ativados em seus valores máximos, há representação do branco. Já se estão com níveis parciais iguais, tem-se como resultado o cinza.

Todavia, a teoria tricromática se mostra incapaz de explicar algumas anomalias na percepção de cores e também outras questões quanto à subjetividade na identificação de cor.

Figura 10 – Absorção da luz por cada tipo de cone em função do comprimento de onda.



Fonte: (GONZALEZ; WOODS, 2006)

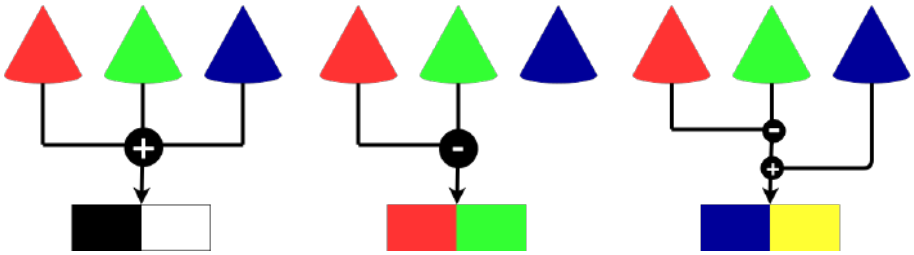
2.2.2 Teoria dos processos oponentes

Desenvolvida por Ewald Hering em 1872, a teoria dos processos oponentes estabelece que a maneira como a percepção de cores ocorre é fundamentada por um sistema de três pares de cores opostas: azul-amarelo, vermelho-verde e preto-branco. Quando um membro do par é ativado, há automaticamente a inibição de atividade no outro. Essa teoria explica que é impossível para o homem perceber combinações de cores do tipo "amarelo azulado"ou "verde avermelhado", já que a ativação ocorre de forma mutuamente exclusiva dentro de cada par (KOSCHAN; ABIDI, 2008).

Através da teoria dos processos oponentes é possível entender, por exemplo, que a percepção do amarelo aconteça ainda que não exista um cone correspondente. A Figura 11 ilustra os pares de cores opostas com suas devidas conexões e ajuda a visualizar que, no caso da produção do amarelo, os três tipos de cones são combinados. Se os cones detectores do vermelho e verde estão ativados, há inibição do azul produzindo assim o amarelo. Por outro lado, a ativação do azul implica na exibição íntegra deste.

Inicialmente, colocadas como concorrentes, as teorias tricromáticas e a de processos oponentes são vistas atualmente como compatíveis se utilizadas para explicação de níveis diferentes do processo visual. Sendo a primeira voltada ao estudo dos estímulos provocados pelas ondas de luz em contato com a retina e o trabalho dos três cones na detecção da cor. Já a segunda teoria fornece um modelo de como os mecanismos que recebem os sinais desses três tipos de cones processam suas mensagens.

Figura 11 – Teoria dos processos oponentes.



Fonte: O autor

2.3 CONSTÂNCIA DE COR

A constância de cor busca eliminar a influência das variações da luz nas cores de uma imagem. Essa invariância permite que características de objetos da cena sejam mais constantes, facilitando assim certas tarefas como de classificação e segmentação (KOSCHAN; ABIDI, 2008).

Uma das formas para conseguir tal invariância é estimar a cor da fonte de luz que incide na imagem e posteriormente usar esse valor para transformar a mesma, de modo que pareça que esta foi capturada sob sob luz canônica. Geralmente, a luz perfeitamente branca $(255, 255, 255)$ é adotada como canônica (GIJSENIJ; GEVERS; WEIJER, 2011).

Matematicamente, uma imagem colorida E pode ser definida como o seguinte produto (BIANCO; CUSANO; SCHETTINI, 2015):

$$E_k(x, y) = \int_\omega I(x, y, \lambda) R(x, y, \lambda)\, \mathbf{C}_k(\lambda) d\lambda, \tag{2.1}$$

onde k representa cada canal de cor (com $k \in \{r, g, b\}$), $I(x, y, \lambda)$ é a radiância da fonte de luz; $R(x, y, \lambda)$, a reflectância da superfície; $\mathbf{C}_k(\lambda)$, as sensibilidades espectrais do sensor em função do comprimento λ de onda da luz sobre o espectro visível ω.

Em sua grande maioria, os métodos para constância de cor adotam algumas suposições sobre a imagem, com isso os esforços são direcionados para descontar o efeito da iluminação. Por exemplo, se for considerado que a reflectância $R(x, y, \lambda)$ é constante e que a função de sensibilidade do sensor $\mathbf{C}(\lambda)$ é conhecida, então qualquer variação na iluminação $I(\lambda)$ causará mudança na cor da imagem (AGARWAL et al., 2006). Outra suposição fortemente adotada é quanto à uniformidade da cor da fonte de luz, fazendo com que a intensidade da mesma seja constante ao longo de toda imagem.

O objetivo então da constância de cor é estimar a cor do iluminante da cena $\mathbf{I}(x, y)$, o qual depende da projeção de $I(x, y, \lambda)$ no sensor $\mathbf{C}(\lambda)$:

$$\mathbf{I}(x, y) = \int I(x, y, \lambda)\, \mathbf{C}(\lambda) d\lambda. \tag{2.2}$$

A Figura 12a ilustra o caso em que a imagem foi capturada sob iluminação levemente amarelada. O resultado do processo de balanceamento de cores é exibido na Figura 12b.

É possível perceber que, na versão corrigida, alguns detalhes da cena (como as etiquetas dos disquetes) apresentam cores mais próximas daquelas que seriam esperadas se a cor do iluminante fosse perfeitamente branca.

Figura 12 – Exemplo de correção de cor.

(a) imagem original (b) versão corrigida

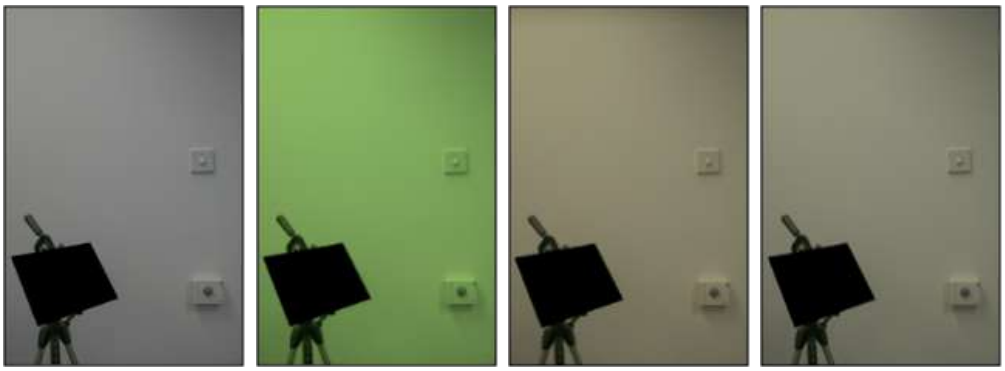
Fonte: O autor

A fim de calcular o erro na estimativa de iluminação, Hordley e Finlayson (2004) propõem uma medida baseada no ângulo formado no espaço de cores RGB entre o iluminante estimado e_b e o real e_a. Nesse sentido, quanto menor o ângulo retornado pela medida, menor o erro associado à estimativa. Por usar o erro angular (definido na Equação 2.3), que independe da intensidade da iluminação, se tem maior facilidade na comparação entre algoritmos sobre um conjunto de imagens.

$$erro_{angular}(e_a, e_b) = \frac{180}{\pi} \cos^{-1}\left(\frac{\parallel e_a \cdot e_b \parallel}{\parallel e_a \parallel \parallel e_b \parallel}\right). \tag{2.3}$$

Um exemplo do que o erro angular pode representar numa variação de iluminante em uma cena pode ser visto na Figura 13. Percebe-se que a cena com cores mais próximas ao ideal (ver Fig. 13d) apresenta menor valor de erro.

Figura 13 – Exemplo de variação do erro angular.



(a) *ground truth* (b) imagem de entrada (c) erro angular= $7,80°$ (d) erro angular= $2,54°$

Fonte: O autor

2.4 MAPAS DE SALIÊNCIA

Ao observar uma imagem, o ser humano é capaz de rapidamente detectar visualmente regiões distintas (chamadas de salientes) que se destacam em relação ao plano de fundo da cena. Essa habilidade vem ganhando cada vez mais a atenção na área de visão computacional, uma vez que, ao ser modelada, pode atuar na etapa de pré-processamento em sistemas computacionais como o de reconhecimentos de objetos, por exemplo. Assim fizeram Alexe, Deselaers e Ferrari (2010), os quais aplicaram um algoritmo de saliência para descobrir regiões que poderiam ser um objeto, repassando posteriormente tais regiões para outros algoritmos mais específicos (detectores de face, de letras e palavras, entre outros).

O mapa de saliência consiste em uma mapeamento onde a cada pixel da imagem é atribuída uma importância relativa, que idealmente estará concentrada em uma região ou objeto mais importante. São inspirados no comportamento do cérebro humano que possui um mecanismo que, de forma pré-atentiva, seleciona as regiões mais interessantes de uma cena para focar a atenção do observador nelas (PAULA et al., 2015).

Existem dois tipos de abordagens para esse tipo de mapeamento: a *bottom-up* e a *top-down*. A metodologia *bottom-up* faz a detecção da região saliente baseada em características intrínsecas à imagem como a cor, contraste ou orientação. Já abordagens do tipo *top-down* utilizam alguns conhecimentos prévios (como o contexto da cena) para realizarem um busca por objetos já conhecidos e esperados em situações específicas como, por exemplo, a detecção de carros em avenidas. Este último método, por ser considerado mais complexo de se modelar, acaba sendo menos adotado em relação à abordagem *bottom-up*(BORJI et al., 2014).

Seguindo a metodologia *bottom-up*, Hou e Zhang (2007) apresenta o algoritmo de saliência SR (sigla do inglês *Spectral Residual*). Esse método propõe um modelo de espectro residual o qual, ao receber o espectro logarítmico da imagem, subtrai deste padrões previamente conhecidos correspondentes ao *background* da cena. O mapa de saliência é, então, obtido aplicando-se a Transformada Inversa de Fourier Sobre o espectro residual resultante. No entanto, quando testado esse método retornou mapas de saliência, geralmente, com aspectos bem abstratos, um exemplo é exibido na Figura 14.

Em (LI et al., 2013) é proposto o algoritmo DSR (sigla do inglês *Dense and Sparse Reconstruction*) voltado para detecção de objetos salientes baseado em erros de reconstrução. Primeiramente, as bordas da imagem são extraídas por meio de superpixels e consideradas como modelos prováveis para plano de fundo. Após isso, o restante imagem é submetido a dois tipos de reconstrução: a Esparsa e a Densa. Os erros gerados por essas reconstruções são combinados através de uma integração Bayesiana para geração de um único mapa de saliência. A Figura 15 traz um exemplo da aplicação desse método.

Já o modelo de saliência GR proposto por Yang, Zhang e Lu (2013) aposta nas técnicas de fecho convexo e regularização de grafo para destacar uniformemente o objeto de interesse e suprimir simultaneamente o plano de fundo efetivamente. Inicialmente, calcula-se

Figura 14 – Aplicação de método SR.

(a) imagem de entrada (b) mapa de saliência

Fonte: O autor

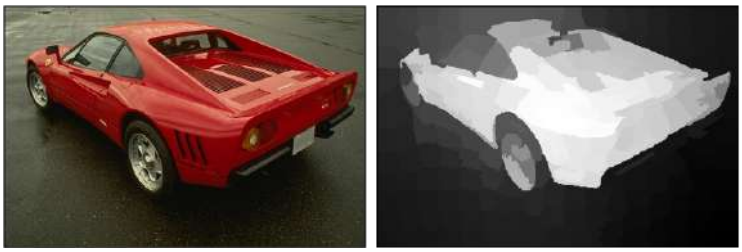
Figura 15 – Aplicação de método DSR.

(a) imagem de entrada (b) mapa de saliência

Fonte: O autor

um mapa de saliência utilizando informações de contraste. No entanto, tal mapa geralmente detecta incorretamente alguns superpixels do *background* sendo necessária uma etapa de refinamento. Para que esse refinamento possa ocorrer, calcula-se antes o centro da região saliente através da técnica de fecho convexo, só então, o aprimoramento do mapa é feito através da regularização de grafo que incentiva os pixels ou segmentos adjacentes a terem o mesmo valor de saliência (ou suavidade). A Figura 16 ilustra um caso de aplicação do método GR.

Figura 16 – Aplicação de método GR.



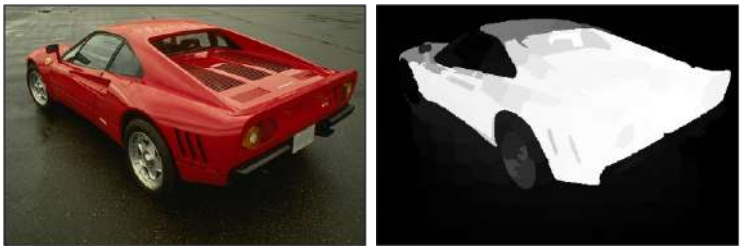
(a) imagem de entrada (b) mapa de saliência

Fonte: O autor

Por sua vez, Jiang et al. (2013) propõem o uso de Cadeias Absorventes de Markov para a detecção de regiões salientes. A imagem é transformada em uma cadeia de Mar-

kov composta por superpixels que podem ser nós de absorção ou de objetos, sendo os primeiros atribuídos aos superpixels das regiões de borda da cena. Esse algoritmo tem como suposição que as regiões de interesse estão afastadas das quatro bordas da imagem, consideradas como pertencentes ao plano de fundo. Aplicando-se caminhadas aleatórias, é calculado para cada nó o tempo de absorção, ou seja, o tempo necessário para chegar aos superpixels de *background*. A ideia é quanto maior for esse tempo, maior deve ser o valor do pixel no mapa de saliência. A Figura 17 ilustra um caso de aplicação do método que neste trabalho é chamado de MC.

Figura 17 – Aplicação de método MC.



(a) imagem de entrada (b) mapa de saliência

Fonte: O autor

2.5 APRENDIZAGEM PROFUNDA

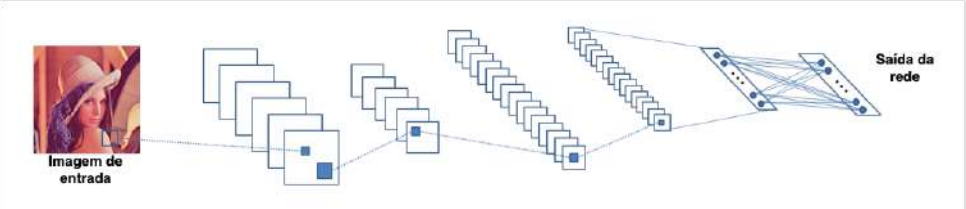
Estudos sobre o córtex visual humano mostram que quando a retina recebe algum estímulo visual, o mesmo é codificado sob forma de sinal e percorre uma sequência de regiões do cérebro. Por sua vez, cada região é responsável pela extração de características específicas que, quando combinadas com as outras anteriormente identificadas, resultam em traços mais complexos (DICARLO; ZOCCOLAN; RUST, 2012). Geralmente, esse processo é repetido até que haja detecção de atributos mais abstratos como faces ou objetos específicos (BENGIO; COURVILLE; VINCENT, 2013). De forma similar, a aprendizagem profunda emprega técnicas que tentam realizar o aprendizado de características de forma hierárquica, fazendo com que as de maior nível de abstração sejam resultados das combinações de atributos de mais baixo nível (SCHMIDHUBER, 2015).

A Aprendizagem Profunda é uma subárea de Aprendizagem de Máquina e consiste em uma classe de algoritmos que objetivam a criação de representações hierárquicas de alto nível dos dados de entrada. O funcionamento desses algoritmos envolve o uso de várias camadas compostas de processamentos não-lineares para extração de características. A ideia é que cada camada use a saída daquela que a precede como entrada, repassando para as demais características mais complexas.

Existem diversas arquiteturas de aprendizagem profunda usadas para os mais diversos fins, incluindo reconhecimento de objeto, classificação e segmentação de imagens. Entre

elas, estão as redes neurais convolucionais (*convolutional neural networks*, CNN) que, desde sua origem, têm se mostrado adequadas para reconhecimento de padrões em imagens ou até outras aplicações que recebem informações 2D como entrada. A Figura 18 exemplifica a arquitetura típica de uma rede convolucional.

Figura 18 – Possível arquitetura de uma CNN.



Fonte: Adaptado de (BEZERRA, 2016)

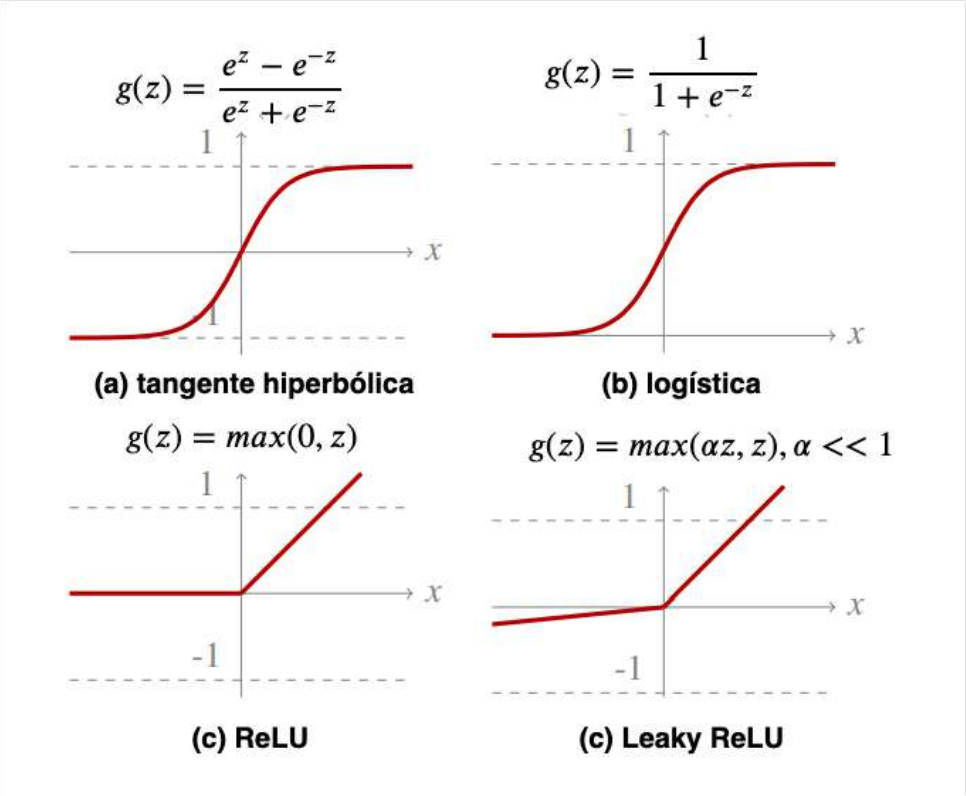
Dentro de uma CNN ocorrem vários processos de extração de características baseados na aplicação de filtros na imagem de entrada, essas operações são chamadas de convoluções. Os filtros utilizados pelas camadas convolucionais se comportam como pequenas janelas que, ao deslizarem sobre a imagem, realizam a extração de características em cada região. Conforme a informação de entrada vai sendo processada pela rede, cada camada convolucional vai se especializando na detecção de atributos por nível. Sendo as primeiras voltadas para percepção de aspectos mais básicos (bordas, por exemplo) e as últimas, de formas mais complexas (curvas, cantos, etc).

Aplicada às camadas convolucionais está a função de ativação que aplica transformação não-linear sobre os dados de entrada. Essa tipo de função permite ao sistema resolver problemas mais complexos que exigem mais do simples transformações lineares. Entre os mais variados tipos de função de ativação, estão a sigmoide logística, a tangente hiperbólica, a retificada linear (*rectified linear units*, ReLU) e a *Leaky* ReLU. Tanto as definições matemáticas quanto as representações gráficas dessas funções estão esquematizadas na Figura 19.

As duas primeiras funções são mais comuns em redes neurais, porém saturam a partir de um determinado ponto. Já a ReLU funciona como uma função identidade para valores positivos, no entanto para valores negativos a saída é sempre zero comprometendo, assim, a atualização dos pesos da rede. A fim de contornar o problema enfrentado pela ReLU, a *Leaky* ReLU insere um hiper parâmetro α com valor pequeno, o qual faz com que a derivada na região negativa da curva ainda seja positiva. Para redes profundas, as funções ReLU e *Leaky* ReLU têm recebido maior adesão por parte da comunidade científica (NAIR; HINTON, 2010).

Além das camadas convolucionais, uma CNN também possui camadas *pooling*. Estas são responsáveis por simplificar a informação da camada anterior, reduzindo as dimensões do conjunto de entrada sem alterar significativamente o seu aspecto (a profundidade do volume é preservada). Existem várias formas pelas quais essa redução pode ser feita.

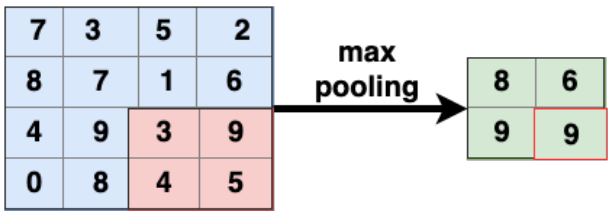
Figura 19 – Funções de ativação.



Fonte: Adaptado de (PONTI; COSTA, 2018)

Algumas envolvem, por exemplo, a escolha da média (*average pooling*) ou do valor máximo (*max pooling*) do conjunto. Além de reduzirem o custo computacional, as camadas *pooling* tendem a tornar a CNN mais independente da posição em que o objeto pode aparecer.

Figura 20 – Exemplo de operação *max pooling* com filtro de tamanho 2×2 e tamanho do passo igual a 2.



Fonte: O autor

Trabalhos recentes têm usado o poder das redes neurais convolucionais para aprender modelos capazes de resolver o problema de constância de cor. No próximo capítulo, são abordados alguns desses.

3 METODOLOGIAS PARA APRIMORAMENTO DE IMAGENS

3.1 METODOLOGIAS PARA CONSTÂNCIA DE COR

Diversos métodos foram propostos para constância de cor, entre eles estão os considerados clássicos na literatura, como *white patch retinex*, *gray world* e a abordagem baseada em mapeamentos *gamute*. Recentemente também vêm surgindo os que aplicam aprendizagem de máquina para correção das cores, alguns dos quais são apresentados ao final desta seção.

Os algoritmos *white patch retinex* e *gray world* são considerados simples e tradicionais, no entanto, ainda são amplamente utilizados como modelos para comparações (AGARWAL et al., 2006). Ambos os métodos assumem que a iluminação é uniforme em toda imagem, ou seja, a irradiância não depende das coordenadas (x, y).

3.1.1 White patch retinex

A ideia básica do *white patch retinex*, apresentado em (LAND, 1977), é que a região mais clara na cena reflete a luz máxima possível para cada canal de cor. A detecção dessa reflectância máxima é importante, pois a cor da mesma indicaria a cor da iluminação. Dessa forma, a estratégia desse algoritmo é estimar o iluminante através dos valores máximos presentes em cada canal de cor, como mostra a Equação 3.1:

$$o_k(x, y) = \frac{c_k(x, y)}{L_{i,max}}, \tag{3.1}$$

onde o_k é a cor de saída do pixel no canal k (com $k \in \{r, g, b\}$) nas coordenadas (x,y), c_k é a cor do pixel na imagem de entrada e $L_{k,max}$ representa o valor máximo em cada canal.

Uma desvantagem do *white patch* é que, se a cena possuir alguma região muito clara por causa de um objeto que não reflete a cor do iluminante, então a estimativa não será equivalente à cor do iluminante (EBNER, 2007). Uma forma de contornar essa situação, é computar um histograma H_k para cada canal k e, posteriormente, definir um ponto de corte com base em uma porcentagem p previamente definida (de preferência, adotam-se valores baixos para p, como por exemplo 1%). Tal ponto de corte é estabelecido de forma que pontos muito brilhantes na cena sejam desconsiderados pelo algoritmo no momento de busca pelos máximos em cada canal. No entanto, a suposição de que na cena há uma única fonte de luz pode limitar o desempenho do algoritmo em casos de dois ou mais iluminantes.

3.1.2 Gray world

O método *gray world*, proposto por Buchsbaum (1980), baseia-se na suposição de que a reflectância média em uma cena sob uma fonte de luz neutra é acromática. Dessa forma,

qualquer mudança na cromaticidade média da imagem é causada pelos efeitos da fonte de luz. A cor da fonte de luz é estimada, então, através da média de cada canal de cor da imagem. A Equação 3.2 apresenta a definição do método:

$$o_k(x, y) = c_k(x, y)\left(\frac{f}{a_k}\right) \tag{3.2}$$

onde a_k é a média de cor em cada canal e f a média sobre os valores a_k encontrados.

Na prática, o *gray world* tende a retornar melhores resultados que o *white patch retinex*, uma vez que se baseia na média de um grande número de pixels e não apenas em um único valor máximo. Dessa forma, evita-se que um único pixel muito brilhante leve o algoritmo a uma predição incorreta. Todavia, o *gray world* possui a mesma limitação do algoritmo anterior quanto ao fato de não funcionar bem em casos de múltiplas iluminações (EBNER, 2007).

A Figura 21 traz os resultados visuais e quantitativos obtidos pelos dois algoritmos. Na parte qualitativa, percebe-se que o *white patch* apenas amenizou o aspecto "alaranjado" da imagem, enquanto que o *gray world* deixou a imagem com um tom mais "azulado". Nesse caso, fica como critério subjetivo definir que imagem apresenta melhor resultado (isso, claro, considerando que não se possui a imagem original iluminada por luz branca).

Já para análise quantitativa, o erro angular, definido pela Equação 2.3, foi utilizado para calcular a distância entre o *ground truth* da iluminação e_a e sua estimativa e_b. Nesse exemplo, o *gray world* apresentou uma estimativa mais próxima do iluminante real.

Figura 21 – Exemplos de correção de cor.

(a) Imagem de entrada (b) *W. patch*, erro=$14,59°$ (c) *G. world*, erro=$6,88°$

Fonte: O autor

Inspirado no algoritmo *gray world*, o método *gray edge*, apresentado em (WEIJER; GEVERS; GIJSENIJ, 2007), assume que a média das diferenças de reflectância é acromática em uma cena. Tal hipótese originou-se da observação feita por Weijer, Gevers e Bagdanov (2005), os quais perceberam que a distribuição derivativa de cores de uma imagem tem uma forma regular semelhante a uma elipsoide, da qual o eixo maior coincide com a cor da fonte de luz. Em outras palavras, isso mostra que a essa distribuição tem maior variação na direção da fonte de luz. Assim como os métodos *gray world*, a abordagem também utiliza a norma Minkowski sobre as derivadas para se ter uma aproximação dessa direção, aplicando

para isso uma equação que assume que a norma de Minkowski de p-grau calculada sobre a derivada da reflectância de uma cena é acromática. Uma das vantagens desse método é que não existe a necessidade de uma base de imagens capturadas sob uma fonte de luz conhecida para calibração.

3.1.3 Abordagem baseada em mapeamentos *gamute*

Proposto inicialmente por Forsyth (1990), esse tipo de método adota o pressuposto de que, em imagens do mundo real, para um determinado iluminante, observa-se apenas um número limitado de cores. O conhecimento sobre esse conjunto restrito de cores (chamado de *gamute* canônico) permite que, caso a iluminação sofra alguma alteração na sua cor, tal desvio possa ser calculado, já que um desvio na cor do iluminante provoca na imagem cores diferentes das esperadas. Dada uma iluminação, o conjunto *gamute* canônico referente é encontrado em uma fase de treinamento, onde procura-se observar a cor refletida por diferentes superfícies sob essa fonte de luz (GIJSENIJ; GEVERS; WEIJER, 2011).

Basicamente, o método consiste em três partes:

1. Ao receber uma imagem capturada sob uma iluminação desconhecida, o *gamute* de entrada é calculado considerando todas as cores presentes na cena;
2. Em seguida, é definido um conjunto de mapeamentos responsável por fazer a convergência entre o conjunto *gamute* de entrada para o *gamute* canônico;
3. É definido um critério para seleção de um dos mapeamentos encontrados na etapa anterior, como, por exemplo, as médias ponderadas do conjunto de mapeamentos.

Caso nenhuma solução de mapeamento seja encontrada, o algoritmo não consegue estimar o iluminante. Outra desvantagem é que, comparado a outros algoritmos, esse tipo de abordagem é considerado mais caro computacionalmente (EBNER, 2007).

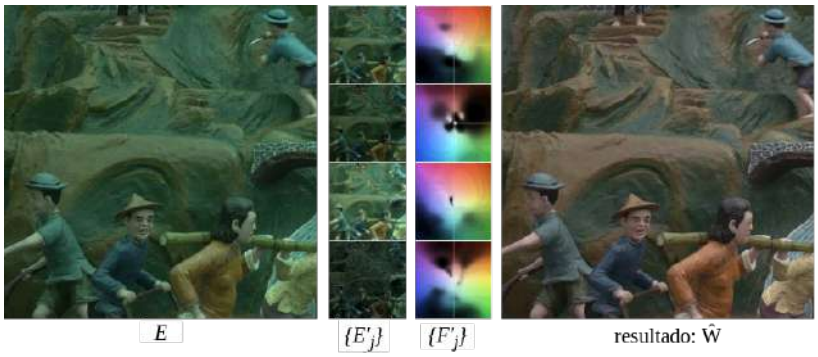
3.1.4 Métodos baseados em aprendizagem de máquina

Os algoritmos que aplicam processos de aprendizagem são caracterizados por fazerem a estimativa da iluminação a partir de um modelo aprendido com base nos dados de treinamento. Inicialmente, foram propostas soluções que utilizam redes neurais (AGARWAL et al., 2006). Essas abordagens, geralmente, possuem entradas recebendo um histograma de cromaticidade da imagem de entrada, enquanto que sua saída consiste de três valores correspondentes à cor do iluminante.

Com a popularização das redes neurais convolucionais e outros modelos de redes profundas, a entrada para os métodos passou a ser a própria imagem, podendo a saída consistir na mesma já corrigida. Tais métodos, quando bem treinados, tendem a retornar melhores resultados. Todavia, também exigem uma grande quantidade de imagens para treinamento (GIJSENIJ; GEVERS; WEIJER, 2011).

Em (BARRON, 2015), foi observado que mudanças na imagem causadas pela cor da iluminação têm impacto direto nos histogramas de cromaticidade da mesma. O problema de constância de cor é reformulado, então, como um problema de localização de um modelo (ou padrão) dentro de um espaço bidimensional, para o qual já se tem técnicas bem conhecidas. A partir de uma imagem de entrada, é gerado um conjunto de imagens com mesma escala, em que cada uma ressalta diferentes aspectos (bordas, texturas, etc) da imagem original. Desse conjunto, são obtidos os histogramas de cromaticidade os quais são convoluídos por uma série de filtros aprendidos sobre uma base de treinamento e depois somados. A cor do iluminante é atribuída ao *bin* do histograma que obtiver maior peso e a imagem de saída é produzida dividindo-se a imagem original pelo iluminante estimado. A Figura 22 traz uma visão geral do método proposto. Apesar dessa abordagem permitir o uso de grandes *datasets* voltados para a área de detecção de objetos, muita informação não é aproveitada. Um exemplo é o contexto semântico que é fortemente ignorado (HU; WANG; LIN, 2019), uma vez que as informações espaciais são minimamente codificadas nos histogramas de cromaticidade. Além disso, assim como *gray world* e *white patch*, esse método também considera a existência de uma única fonte de luz.

Figura 22 – Visão geral da metodologia proposta por Barron (2015). A partir de uma imagem de entrada E, é gerado um conjunto de imagens E'_j onde são destacados alguns aspectos da cena. Sobre E'_j extraem-se os histogramas de cromaticidade que são convoluídos pelos filtros F'_j. O resultado é representado por $\hat{\mathrm{W}}$.



Fonte: Adaptado de (BARRON, 2015)

Por sua vez, ao invés de histogramas, o algoritmo proposto por Lou et al. (2015) submete a própria imagem como entrada de uma CNN com 8 camadas, a fim de estimar a cor da fonte de luz. Nesse método, a fase de pré-treinamento recebeu uma atenção maior, sendo dividida em duas partes. A primeira parte do pré-treinamento consiste em treinar a rede para o problema de classificação, utilizando a base ImageNet (DENG et al., 2009); a saída da mesma é então adaptada para as 1000 categorias existentes no *dataset*. Com isso, obtém-se um modelo pré-treinado com características de imagens em geral.

Para a segunda etapa, um algoritmo já existente de constância de cor é utilizado para

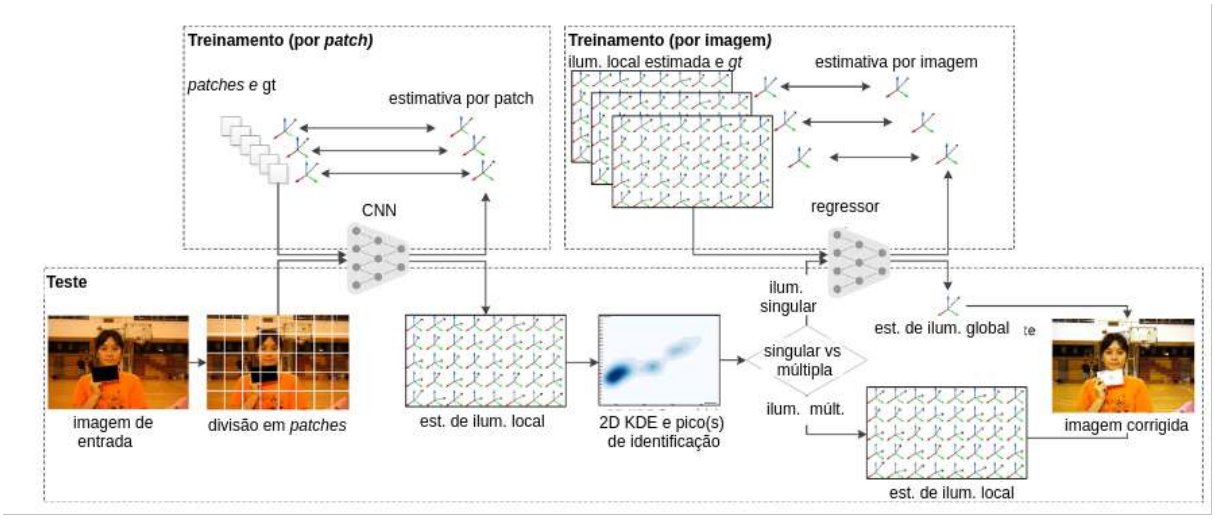
fornecer estimativas quanto à cor da iluminação para cada imagem da base ImageNet. Neste caso, a última camada é alterada de forma a ter três saídas (r,g,b) as quais são repassadas para uma função de distância Euclidiana juntamente com a iluminação estimada pelo algoritmo escolhido. Após a fase de pré-treinamento, segue-se a etapa de treinamento normal com as bases que fornecem o valor real da iluminação. O custo computacional na etapa de pré-processamento pode tornar a aplicação desse método não tão atraente, além do fato de se ter mais uma variável na abordagem que é a escolha do algoritmo para a segunda fase do pré-treinamento.

Adotando como entrada as *patches* de uma imagem, Bianco, Cusano e Schettini (2015) propõem uma CNN para predição da iluminação. A ideia é que, a partir de uma imagem colorida, são geradas *patches* não sobrepostas para as quais o iluminante local será estimado. No final, as estimativas locais são combinadas para obter-se uma para global. Posteriormente, em (BIANCO; CUSANO; SCHETTINI, 2017), os mesmos autores apresentam um método que se adapta também para casos em que há múltiplas iluminações na cena. A Figura 23 mostra alguns processos que ocorrem dentro das fases de treinamento e teste. Ainda recebendo *patches* não sobrepostas, a CNN é responsável por fornecer uma estimativa local para cada porção da imagem submetida. Tais saídas são então analisadas para detectar-se a presença ou não de múltiplas iluminações. Quando vários iluminantes são detectados, as estimativas locais são utilizadas diretamente para a geração da imagem de saída. Caso contrário, o algoritmo aplica um processo de regressão sobre as estimativas locais de forma a retornar um único valor para iluminação, que será utilizado para correção da cor. Dessa forma, Bianco, Cusano e Schettini (2017) conseguem tratar casos de múltiplos iluminantes. Todavia, a rede proposta tem a tendência de se concentrar especificamente em regiões da face (HU; WANG; LIN, 2017), o que é um limitante em situações onde há ausência de tal característica.

Para lidar com *patches* que talvez possam ser ambíguas quanto à iluminação, Shi, Loy e Tang (2016) propõem uma metodologia baseada na geração e seleção das hipóteses mais plausíveis para o iluminante. Para isso, a rede proposta consiste em duas sub-redes onde a primeira, ao receber um *patch*, fornece múltiplas saídas para iluminação. Por sua vez, a segunda sub-rede escolhe apenas a estimativa que for mais próxima da cor real de acordo com uma função de perda Euclidiana. Este método garante ser aplicável em casos de múltiplos iluminantes, porém não fica claro quais aspectos devem ser levados em consideração para a definição da quantidade de hipóteses a serem geradas para cada *patch*. No trabalho em questão, são formuladas duas hipóteses por região; no entanto, tal valor pode não ser suficiente para lidar com imagens com três ou mais fontes de iluminação.

Assim como Barron (2015), Oh e Kim (2017) desenvolvem uma metodologia (Figura 24) que adapta o problema da constância de cor para outro mais conhecido. A ideia é tratar a tarefa de estimar a cor como um problema de classificação, em que cada classe representa um iluminante possível. Para isso, uma etapa de agrupamento é executada previamente

Figura 23 – Esquema proposto por Bianco, Cusano e Schettini (2017) para constância de cor.



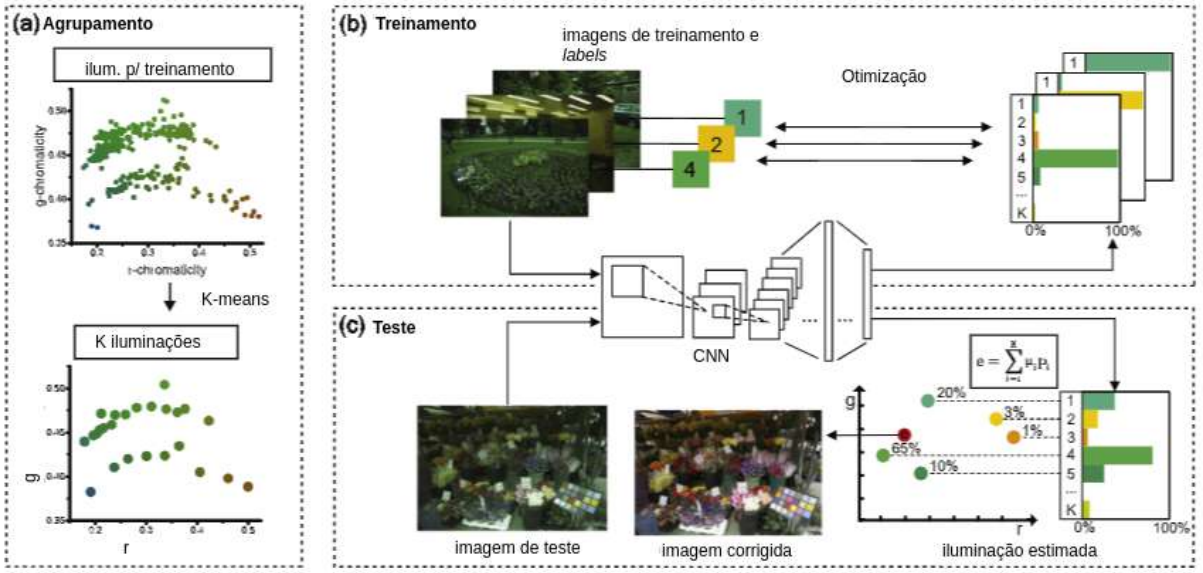
Fonte: Adaptado de (BIANCO; CUSANO; SCHETTINI, 2017)

sobre os iluminantes reais da base de treinamento, o objetivo é agrupar as iluminações que são mais semelhantes em um só grupo, atribuindo a um mesmo rótulo. Dessa forma, pretende-se facilitar a classificação pela rede, pois reduz-se a quantidade de classes a serem reconhecidas. A rede utilizada é uma versão adaptada da AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) e que, depois de pré-treinada com a ImageNet, recebe os *patches* com o rótulo referente à imagem original. As saídas obtidas para cada região da imagem são então combinadas de forma a gerar um único valor para iluminação com o qual se pode fazer a correção de cor.

Uma desvantagem dessa abordagem é que *patches* com pouca informação semântica têm o mesmo peso que as demais para a estimativa global do iluminante. Além disso, a metodologia não considera casos onde existem várias iluminações incidindo na mesma imagem. A Figura 25 traz um exemplo em que esse método é aplicado.

Embora também apliquem uma abordagem que considera todas as estimativas locais para o cálculo da iluminação global, Hu, Wang e Lin (2017) apontam limitações em trabalhos anteriores que utilizam essa estratégia. A crítica levantada consiste no fato de que, nesses trabalhos, todos os *patches* de uma imagem têm suas estimativas tratadas com mesmo peso, solução não tão adequada em casos onde existem regiões da cena que, por possuírem pouco conteúdo semântico, podem dificultar a inferência. Essa dificuldade é exemplificada pela Figura 26, a qual retrata a situação de um *patch* considerado ambíguo, uma vez que, a partir dele, diversas combinações de iluminação e reflectância são possíveis. No caso da parede, isto ocorre, porque, para esta classe, diversas cores podem lhe ser atribuídas (tais como branco, amarelo e outras). Para esse tipo de *patch*, é sugerido que suas estimativas locais sejam ponderadas com menor peso (menor confiança) dentro do

Figura 24 – Esquema proposto por Oh e Kim (2017) para constância de cor.



Fonte: Adaptado de (OH; KIM, 2017)

Figura 25 – Exemplo de aplicação do método proposto por Oh e Kim (2017).

(a) Imagem de entrada (b) resultado. Erro=$0,104°$

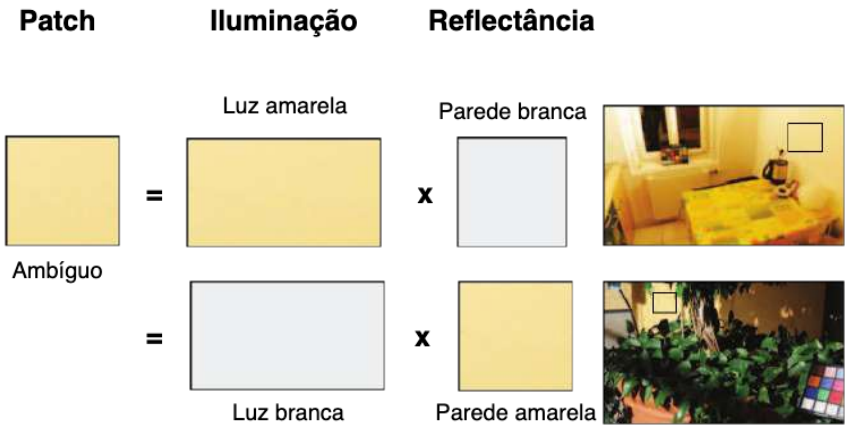
Fonte: O autor

cálculo para iluminação global.

Já para os *patches* informativos, a saber, aqueles que contêm objetos com cores mais inatas (como bananas, maçã, grama, etc) receberiam pesos maiores (maior confiança). A Figura 27 ilustra o caso em que o objeto contido no fragmento da cena possui uma reflectância conhecida (espera-se que a cor da banana seja amarela), diminuindo assim as possibilidades para o iluminante (a hipótese da iluminação ser amarela já é bem menos provável, por exemplo).

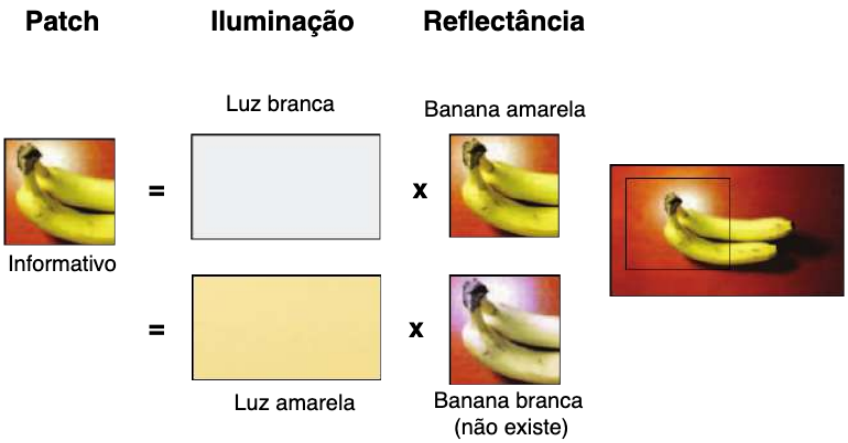
A arquitetura original da rede FC4, apresentada na Figura 52, considera simultaneamente todas as estimativas locais de iluminação. Isto é possível já que nesse tipo de rede todas as convoluções são compartilhadas. Além disso, a arquitetura é capaz de suportar entradas de qualquer tamanho, evitando possíveis distorções causadas pela operação de redimensionamento (LOU et al., 2015).

Figura 26 – Exemplo de *patch* ambíguo.



Fonte: Adaptado de (HU; WANG; LIN, 2017)

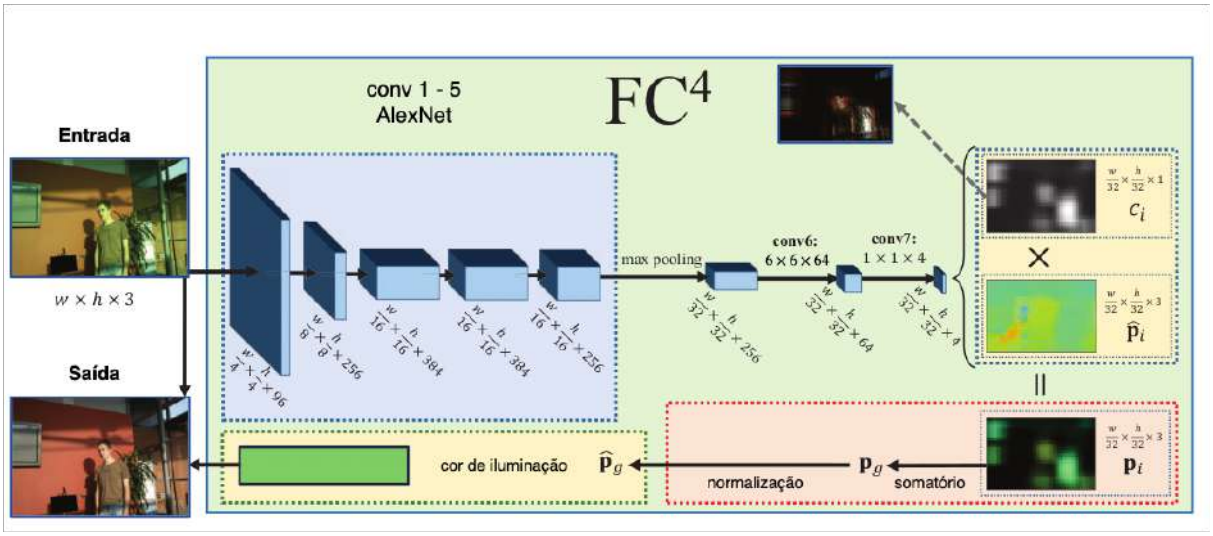
Figura 27 – Exemplo de *patch* informativo.



Fonte: Adaptado de (HU; WANG; LIN, 2017)

Inicialmente, da imagem de entrada pretende-se extrair suas características semânticas. Essa informação é útil à rede, pois ajuda no processo de distinção das regiões que são informativas e das que podem ser ambíguas. Para esta tarefa, foram escolhidas as cinco primeiras camadas convolucionais (conv1-5) de uma AlexNet (KRIZHEVSKY; SUTSKEVER; HINTON, 2012) já pré-treinada sobre a base ImageNet (DENG et al., 2009). Para o contexto de constância de cor, é esperado que modelo de rede não seja invariante às mudanças de iluminação. Esse requisito acaba restringindo a possibilidade de substituição da AlexNet por outras arquiteturas consideradas mais eficazes em problemas de classificação. Visto que tais modelos tendem a extrair informações semânticas da imagem de forma invariante à mudança de iluminação. A fim de reduzir o tamanho do modelo aplicado para essa tarefa, Hu, Wang e Lin (2017) aplicam a SqueezeNet (IANDOLA et al., 2016) v1.1 no lugar da AlexNet, obtendo resultados equiparáveis. Todos os experimentos executados neste trabalho aplicam a SqueezeNet como extrator de características semânticas.

Figura 28 – Arquitetura da rede FC4.



Fonte: Adaptado de (HU; WANG; LIN, 2017)

O resultado dessa extração é, então, submetido a uma camada convolucional (conv6) com *kernel* 6×6 e 64 filtros. Para evitar casos *overfitting*, aplica-se *dropout* de 0,5 para conv6. Por sua vez, na sétima convolução (conv7), há apenas a redução de dimensionalidade quanto à profundidade dos mapas de características, aqui chamados de mapas semi-densos. A saída da conv7 consiste de 4 mapas, dos quais os três primeiros representam as estimativas locais de iluminação g e o último, os pesos (confiança c) para cada região R_i. Por fim, a Equação 3.3 descreve como a combinação das estimativas ponderadas por suas confianças resultam no valor da iluminação global estimada $\widehat{p}_g$.

$$\widehat{p}_g = normaliza\Big(\sum_{i \in R} c(R_i) g(R_i)\Big) \tag{3.3}$$

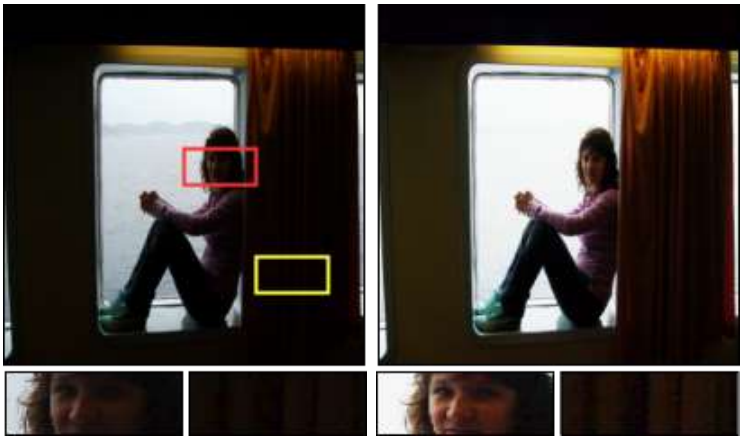
Em (HU; WANG; LIN, 2017), os autores falam sobre uma segunda implementação mais simples e que, teoricamente, levaria a uma precisão semelhante. Nela, as confianças c_i e as estimativas p_i são implicitamente a norma e direção de p_i não sendo, portanto, necessário para conv7 gerar quatro mapas, mas apenas três. Neste trabalho, ambas as configurações para FC4 foram treinadas e testadas separadamente. Para fins de facilitar a referência a cada uma delas, adotou-se a nomenclatura **FC4 V1** para a versão que separa o mapa de confiança das estimativas locais para iluminação. Enquanto que o segundo modelo é chamado de **FC4 V2**.

3.2 METODOLOGIAS PARA IMAGENS FRACAMENTE ILUMINADAS

Em geral, imagens capturadas sob condições de luz insuficientes possuem valores de pixels em um intervalo baixo, o que faz com que a imagem pareça muito escura, podendo dificultar a identificação de objetos ou texturas da cena (LV et al., 2018). A Figura 29 traz

um caso típico de imagem capturada em ambiente pouco iluminado e sua versão aprimorada quanto à iluminação, nesta é possível ver que os detalhes do rosto humano ficam um pouco mais nítidos.

Figura 29 – Exemplo de aprimoramento em imagem fracamente iluminada.



(a) imagem original (b) versão aprimorada

Fonte: O autor.

Em sua maioria, os métodos existentes para esse tipo de aprimoramento podem ser divididos em três grupos: métodos baseados em Equalização de Histograma (HE do inglês *Histogram Equalization*), métodos baseados na teoria Retinex (LAND; MCCANN, 1971) e os que são baseados em inteligência artificial.

3.2.1 Métodos baseados em equalização de histograma

A Equalização de Histogramas é uma técnica que procura melhorar a distribuição das intensidades de uma imagem através de seu histograma. O principal efeito desse método é o aumento do contraste global na imagem, apresentando bons resultados em cenas muito claras ou muito escuras. No entanto, pode deixar a desejar em situações de iluminação não uniforme. Diversos algoritmos baseados em HE tentam contornar essa e outras possíveis limitações da técnica original. Um deles é o CLAHE (sigla do inglês *Contrast Limited Adaptive Histogram Equalization*) primeiramente apresentado em (PIZER, 1990), que implementa uma equalização adaptativa do histograma. A Figura 30a traz o exemplo de uma cena real composta por regiões escuras e pontos de intensa iluminação. Na imagem aprimorada gerada pela técnica CLAHE (Fig. 30b) percebe-se uma certa distorção nas cores dos prédios.

Outro algoritmo HE variacional é o LDR (sigla do inglês *Layered Difference Representation*)(LEE; LEE; KIM, 2013) que foca em aprimorar o contraste da imagem utilizando a representação da diferença de camadas em histogramas 2D para aumentar a diferença de nível de cinza entre pixels adjacentes. Por sua vez, Celik e Tjahjadi (2011) usam informações contextuais inter pixel para construção de histogramas 2D, sobre os quais se busca

Figura 30 – Exemplo de aplicação do método CLAHE (REZA, 2004) em imagem real com fraca iluminação.

(a) Imagem real (b) Versão aprimorada

Fonte: O autor

um mapeamento de forma que o algoritmo concentre atenção nas grandes diferenças de cinza.

3.2.2 Métodos baseados na teoria Retinex

Outra estratégia para o aprimoramento de imagens de baixa iluminação utiliza a teoria Retinex, a qual supõe que a imagem é composta basicamente por dois fatores: reflexão e iluminação. O principal foco dos algoritmos que aplicam tal modelo costuma ser o cálculo da iluminação seguido da remoção de tal fator na cena. Dessa forma, a versão aperfeiçoada da imagem corresponde à reflectância da mesma (JOBSON; RAHMAN; WOODELL, 1997). Entre os trabalhos que aplicam esse paradigma está o de Fu et al. (2016), no qual é proposto o modelo variacional ponderado SRIE (sigla do inglês *Simultaneous Reflectance and Illumination Estimation*) capaz de estimar simultaneamente a iluminação e reflectância de uma determinada imagem. Em termos de funcionalidades, o modelo SRIE é, até onde se sabe, um dos que mais promete correções, pois incluindo também efeitos de constância de cor e supressão de ruído leve. A Figura 31 traz um exemplo, onde uma imagem levemente escura e com ruído branco gaussiano ($\sigma = 5$) é corrigida.

Por sua vez, Guo, Li e Ling (2016) criaram uma abordagem onde o foco consiste em estimar apenas o mapa de iluminação, promovendo assim uma redução no custo computacional. A abordagem, chamada LIME (sigla do inglês *Low-light Image Enhancement*), começa com um mapa de iluminação inicial formado pela busca da intensidade máxima encontrada nos canais R, G e B para cada pixel. Após isso, o mapa é então refinado através da aplicação do método Lagrangiano aumentado, seguido de uma correção Gama. Como pós-processamento, o método LIME inclui, entre outros passos, a passagem do filtro BM3D (DABOV et al., 2008) na reflectância da imagem. No entanto, Tao et al. (2017)

Figura 31 – Exemplo de aplicação do método SRIE em imagem com fraca iluminação e adicionada de ruído gaussiano.

(a) imagem real (b) versão aprimorada

Fonte: Fu et al. (2016)

relata que tal metodologia pode provocar consideráveis distorções de cores (ver Figura 32).

Figura 32 – Exemplo de aplicação do método LIME em imagem real com fraca iluminação.

(a) imagem real (b) versão aprimorada

Fonte: Guo, Li e Ling (2016)

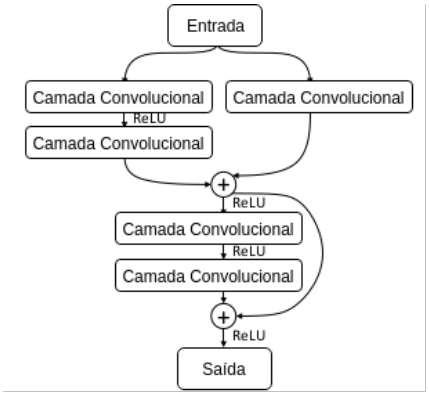
3.2.3 Métodos baseados em inteligência artificial

Um dos primeiros trabalhos, que se tem conhecimento, a apresentar uma abordagem baseada em redes neurais para correção de imagens com pouca luz é o de Lore, Akintayo e Sarkar (2017). Nesse trabalho, um *autoencoder* é construído para aprender a função de clareamento em imagens com pouca luz, fazendo as devidas correções, evitando a saturação partes mais claras em cenas com alta variação de iluminação. Para o treinamento das

redes neurais (*enconder* e *decoder*), exemplos de treinamento sinteticamente escurecidos e acrescentados de ruído foram usados, a fim de tornar o sistema capaz de não somente tratar imagens mal iluminadas, mas também aquelas degradadas, por exemplo, por hardware. No entanto, todas as imagens utilizadas pelos autores são em níveis de cinza o que limita a compreensão do desempenho da rede em imagens coloridas, por exemplo, se há distorções de cor no resultado final, etc.

Utilizando também abordagem de aprendizagem profunda, Tao et al. (2017) desenvolveram uma rede neural convolucional composta por módulos convolucionais que trabalham com mapas de características em diversas escalas. Inspirados em módulos do tipo *inception*, a arquitetura desses módulos pode ser vista na Figura 33, onde também observa-se o uso de aprendizado residual pela presença de *shortcuts*. Quanto à parte experimental do trabalho, percebeu-se a falta de mais testes em imagens naturalmente escuras, sendo a maioria dos experimentos executados em ambientes de iluminação controlada. A Figura 34 traz o único caso de aplicação em imagem não sintética. Nesse caso, percebe-se que a versão resultante (imagem b dessa mesma figura) ficou com baixo contraste e ganhou um aspecto acinzentado.

Figura 33 – Módulo convolucional usado em (TAO et al., 2017)



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

Assim como em (GUO; LI; LING, 2016), Li et al. (2018) também propuseram o uso de mapa de iluminação a ser usado posteriormente para restauração da imagem. No entanto, esse mapa é estimado por uma CNN treinada com imagens escurecidas sinteticamente, geradas com base no modelo Retinex. Uma das limitações reportada é a limitação da rede quanto ao processamento de imagens de baixa qualidade, seja por presença de ruído e/ou por causa de compressão JPEG. O desempenho da CNN é avaliado também em imagens naturais de pouca luz através de um estudo de usuário com grupo de apenas 7 participantes. Alguns resultados desse método são exibidos na Figura 35. Ao analisar as versões aprimoradas (retiradas diretamente do artigo em questão), percebe-se a adição de detalhes que fogem da realidade bem como eventuais distorções de cores. Por exemplo, na imagem noturna (linha superior da Figura 35) foram adicionados reflexos de lâmpadas

Figura 34 – Exemplo de aplicação do método de Tao et al. (2017) em imagem real com fraca iluminação.

(a) Imagem real (b) Imagem resultante

Fonte: (TAO et al., 2017)

na água que fisicamente seria impossível ter. Por sua vez, a cena do trânsito recebeu, em determinados trechos, cores diferentes da cor comum que se espera de uma pavimentação asfáltica.

Figura 35 – Exemplo de aplicação do método de Li et al. (2018) em imagem real com fraca iluminação.

(a) Imagem real (b) Versão aprimorada

Fonte: (TAO et al., 2017)

4 ALGORITMOS PROPOSTOS

4.1 MODIFICAÇÕES PROPOSTAS AO ALGORITMO DE HU, WANG E LIN

De todos os modelos de rede para constância de cor vistos no capítulo anterior, a rede totalmente convolucional (FC4, do inglês: *Fully Convolutional Color Constancy with Confidence-weighted Pooling*), proposta em (HU; WANG; LIN, 2017), é a que retorna as menores médias de erro angular em bases padrões usadas para *benchmarking*, como a de Gehler-Shi (SHI, 2000) e a NUS-8(CHENG; PRASAD; BROWN, 2014).

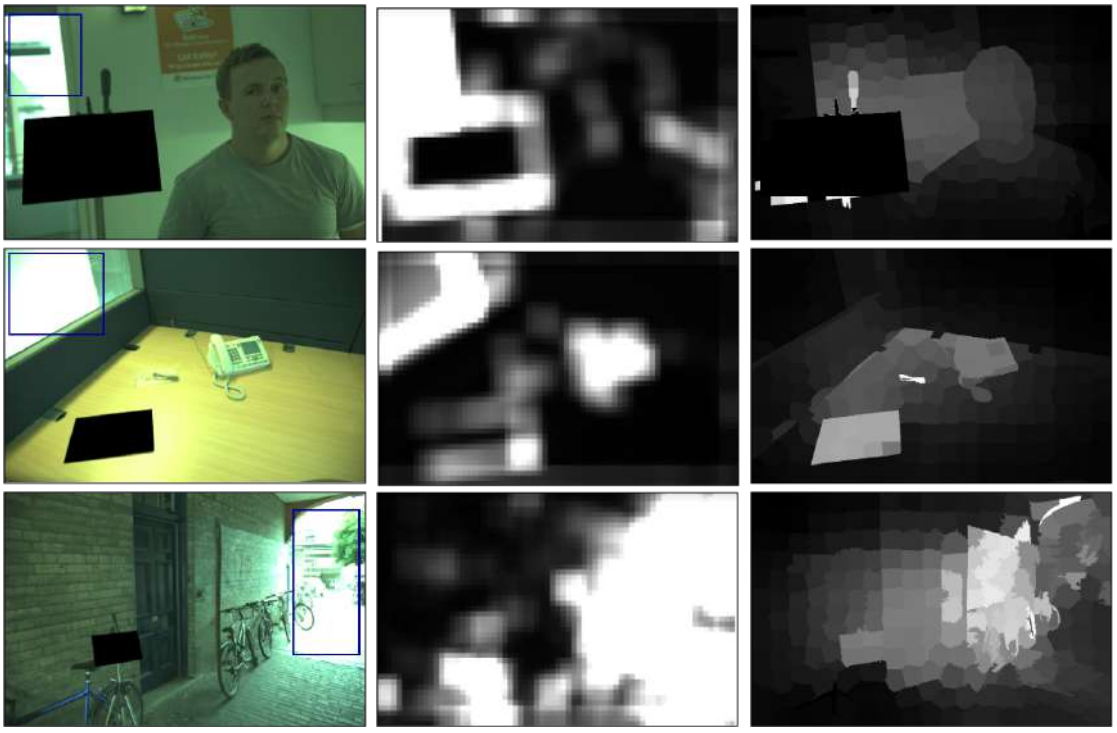
A análise dos mapas de confiança aprendidos pela rede FC4 permite identificar alguns padrões que normalmente recebem maior peso. Regiões como face humana, locais de intenso brilho ou com textura não-lisa são exemplos. No entanto, percebeu-se que os casos em que a rede mais teve dificuldade em estimar o iluminante costumam apresentar regiões com saída de luz intensa. Alguns exemplos desses casos são apresentados pela Figura 36a com seus respectivos mapas de confiança (Figura 36b) e de saliência (Figura 36c). Nos dois tipos de mapeamento, as regiões de maior confiança são representadas em tons de cinza, de forma que quanto maior for o peso para região, mais clara a mesma será. Os mapas de confiança gerados pela rede FC4 indicam a tendência da mesma em concentrar as maiores confianças em locais como janelas e luminárias, chegando até mesmo a ignorar regiões consideradas informativas (como face). Por outro lado, os mapas de saliência demonstram uma melhor distribuição dos pesos entre as regiões. Embora também destaquem *patches* brilhantes.

Dado esse cenário, propôs-se a inclusão de mapas de saliência como mecanismo para ponderação de *patches*, uma vez que tais métodos mostram não ser totalmente atraídos por regiões de forte brilho. Além disso, por retornarem maiores pesos para locais de faces humanas e texturas não uniformes, os mapas de saliência permitem que a rede não deixe de se beneficiar desses elementos quando presentes na cena.

Para a escolha do método detector de regiões salientes, foram testados alguns dos mais promissores segundo o estudo de Borji et al. (2014) (ver capítulo Experimentos para mais detalhes). Destes, o algoritmo MC (JIANG et al., 2013) mostrou boa resposta às regiões de muita luz, não se concentrando apenas nelas como mostram os mapas da Figura 36c. Além disso, para imagens onde a rede FC4 não deu pesos significativos às classes consideradas ricas em características semânticas (faces, por exemplo), os mapas de saliência assim o fizeram. Um exemplo disso está na primeira cena da Figura 36a onde o rosto, praticamente ignorado pela rede, recebe maior confiança pelo método MC.

A primeira modificação proposta, chamada de **mFC4 V1**, consiste em substituir o quarto mapa gerado em conv7 pelo mapa de saliência correspondente à imagem de entrada, sendo este último agora o responsável por atribuir à cada estimativa local $\widehat{p}_i$ a

Figura 36 – Exemplos de imagens com regiões de forte brilho.

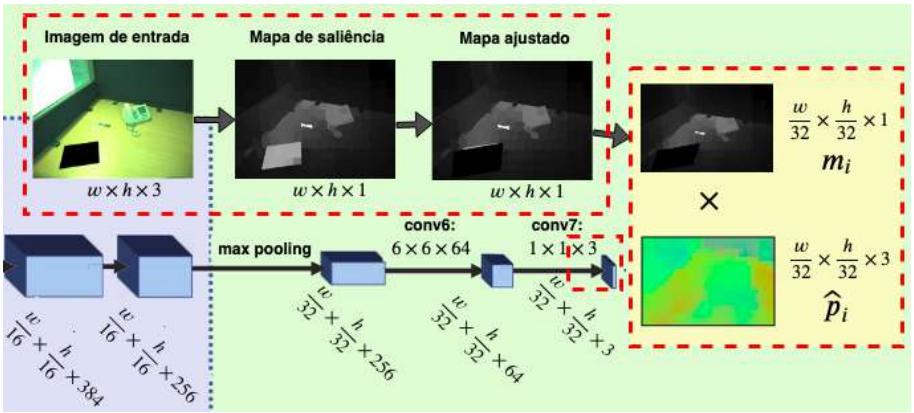


(a) Imagem de entrada (b) Mapa de confiança (c) Mapa de saliência

Fonte: O autor

confiança m_i. Um dos principais objetivos dessa modificação é verificar o desempenho da rede quando a ela está incumbida apenas da tarefa de calcular os iluminantes por *patch*. A Figura 37 traz um recorte feito sobre a arquitetura original (exibido na Figura 52) com a inclusão das pré-etapas (geração e ajuste de mapas), bem como com as alterações quanto à configuração da rede.

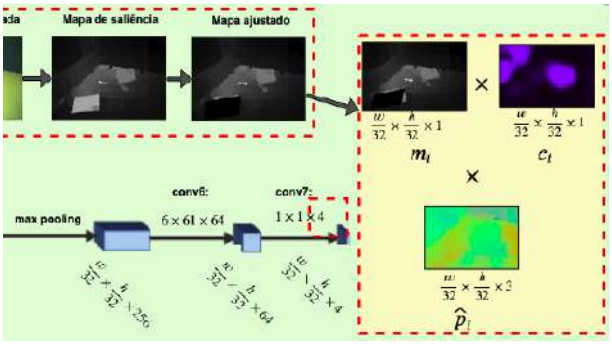
Figura 37 – Primeira modificação proposta: mFC4 V1.



Antes do início de treinamento, aplica-se o detector de saliência MC em todas as imagens para que sejam gerados os mapas. Como em alguns casos a região do *Color Checker* (tabela utilizada para cálculo da cor do iluminante) é destacada, aplica-se sempre uma

máscara com pixels de valor 0 em tais locais, já que estes não devem exercer influência no processo de aprendizagem. Após esse ajuste, o mapa é redimensionado (por transformação bilinear) para as mesmas dimensões dos três mapas de estimativas locais gerados pela conv7. Da mesma forma como no método original, os mapas $\widehat{p}$ passam por uma camada do tipo ReLU a fim de se evitar valores negativos. Após a multiplicação das iluminações regionais $\widehat{p}_i$ pelas suas respectivas confianças m_i, as estimativas locais são somadas e normalizadas, gerando o valor do iluminante global $\widehat{p}_g$. Por fim, para se obter a imagem corrigida, basta dividir-se a imagem de entrada E por $\widehat{p}_g$.

Figura 38 – Segunda modificação proposta: mFC4 V2.



A segunda forma de aplicação dos mapas de saliência na FC4 foi multiplicando-os pelas confianças c_i (ver Figura 38). Dessa forma, para uma região receber alta confiança precisa ser bem pontuada nos dois mapas (mapa de saliência e mapa de confiança). Nos casos onde há saída de luz intensa, tal combinação pode ser útil, já que o peso do mapa de saliência tem a tendência de ser pequeno nessas situações. O resultado da multiplicação dos dois mapas é, então, repassado para a função *softmax*, responsável por fazer a distribuição de probabilidades dentro do conjunto de entrada.

No capítulo de experimentos são apresentados e discutidos os impactos de cada modificação proposta no desempenho da rede.

4.2 ALGORITMO PARA APRIMORAMENTO DE IMAGENS FRACAMENTE ILUMINADAS

A abordagem proposta para o aprimoramento de imagens fracamente iluminadas neste trabalho faz uso da técnica de aprendizagem profunda para estimar a iluminação presente em uma cena, utilizando essa informação para correção da iluminação. No caso, a CNN utilizada é uma adaptação de uma rede neural convolucional aplicada para diminuição de ruído branco gaussiano (DnCNN sigla do inglês *Denoising Convolutional Neural Network*) apresentada por Zhang et al. (2017). Na DnCNN, o ruído é primeiramente estimado por um conjunto de blocos convolucionais e representado, no final, com as mesmas dimensões (comprimento e largura) da imagem de entrada. Como a natureza do ruído em questão é

aditiva, a imagem residual estimada pela rede é subtraída da imagem de entrada ruidosa, gerando uma versão mais limpa da mesma.

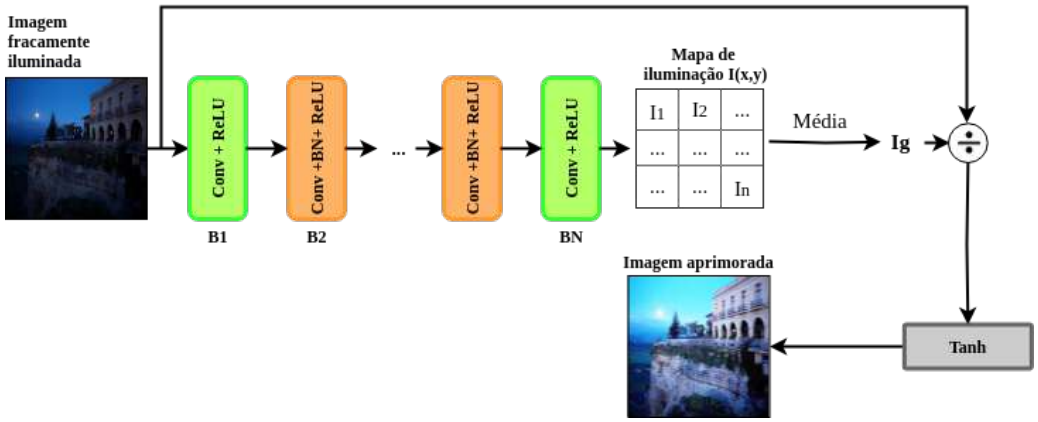
De forma análoga, a CNN proposta neste trabalho também usa uma conexão direta com a entrada para que, em combinação com a iluminação estimada, haja a geração de uma imagem mais iluminada.

Inspirada na rede DnCNN, a adaptação feita para o problema de iluminação é mais visível nas últimas camadas da rede WIIEN (sigla do inglês *Weakly Illuminated Image Enhancement Network*), bem como na função de perda utilizada pela rede. Essas alterações foram feitas com vista à Equação 4.1 que traz o modelo simplificado para formação de imagens:

$$E(x, y) = R(x, y) \cdot I(x, y), \tag{4.1}$$

onde x e y representam as coordenadas em uma imagem E fracamente iluminada, R representa a reflectância e I é a iluminação. Assim como as abordagens baseadas na teoria Retinex, aqui considera-se que imagem aprimorada equivale à reflectância. O mesmo modelo de formação de imagem é usado em (LI et al., 2018), porém a estrutura da rede se limita apenas em estimar a iluminação. Diferentemente do trabalho anterior, a modelagem da rede WIIEN (apresentada na Figura 39) inclui um componente que faz a divisão da imagem de entrada pela iluminação estimada. Desta forma, ao final da rede já se obtém a imagem aprimorada.

Figura 39 – Arquitetura da rede WIIEN.



A arquitetura da rede WIIEN tem por base a da rede DnCNN, proposta em (ZHANG et al., 2017). Voltada para diminuição de ruído gaussiano branco aditivo, a DnCNN foca em aprender a imagem residual, a qual é subtraída da imagem de entrada, gerando uma versão mais limpa da mesma ao final da rede.

Para o contexto de aprimoramento de imagens com pouca iluminação, algumas adaptações foram feitas com vista à Equação 4.1, sendo a principal delas a *skip connection* que aplica divisão. O resultado de tais ajustes é apresentado na Figura 39 onde Conv, BN e

ReLu indicam as camadas de convolução, normalização e função de ativação, respectivamente. Assim como em (ZHANG et al., 2017), a imagem de entrada deve passar por blocos de dois tipos (responsáveis pela não-linearidade): Conv+ReLU e Conv+BN+ReLU, ambos aplicam preenchimento com zeros a fim de garantir que a imagem resultante tenha as mesmas dimensões que a de entrada. Cada camada convolucional (com exceção da última) consiste em 64 filtros com *kernel* de tamanho 3×3 e *stride* 1. A Tabela 1 fornece, de forma mais detalhada, as configurações da rede WIIEN.

Por sua vez, a última convolução é responsável por gerar o mapa de iluminação $I(x, y)$ sobre o qual aplica-se uma redução por média. Essa média, referente às intensidades presentes em $I(x, y)$, é chamada de I_g e é usada por uma *skip connection* para dividir a imagem de entrada. Em seguida, o resultado da divisão passa pela função da tangente hiperbólica que finalmente retorna a imagem aprimorada.

Tabela 1 – Configuração da rede WIIEN

Bloco	**Tipo**	**Entrada**	**Saída**
B1	Conv	$3 \times 40 \times 40$	64
	ReLU	$64 \times 40 \times 40$	64
B2...B16	Conv	$64 \times 40 \times 40$	64
	BN	$64 \times 40 \times 40$	64
	ReLU	$64 \times 40 \times 40$	64
B17	Conv	$64 \times 40 \times 40$	1
	ReLU	$64 \times 40 \times 40$	1

Uma vez que a avaliação da qualidade da imagem está relacionada ao Sistema Visual Humano, não existe uma medida universal que satisfaça subjetivamente os desempenhos de algoritmos para aprimoramento de imagens (WANG et al., 2013). O que existe são medidas que analisam características importantes de uma imagem, como por exemplo, o seu contraste.

Seguindo trabalhos anteriores Tao et al. (2017) e Lv et al. (2018), a função perda utilizada para o treinamento da rede é a SSIM (sigla para *Structural Similarity Index*)(WANG et al., 2004). Aplicada para medir a semelhança entre duas imagens, a medida SSIM faz essa tarefa com base em três tipos de comparação: a de luminância, contraste e estrutural. A Equação 4.2 traz a sua definição:

$$SSIM(w, z) = \frac{2\mu_w\mu_z + C_1}{\mu_w^2 + \mu_z^2 + C_1} \cdot \frac{2\sigma_{wz} + C_2}{\sigma_w^2 + \sigma_z^2 + C_2}, \tag{4.2}$$

onde w e z representam dois sinais de imagem não negativos correspondentes entre si (por exemplo, dois *patches* extraídos de cada imagem), μ_w e μ_z são as médias dos valores dos

pixels sobre cada *patch*, σ_w^2 e σ_z^2 representam as variâncias e σ_{wz} é a covariância. Para evitar que o denominador fique igual a zero, são utilizadas as constantes C_1 e C_2. O valor SSIM retornado é um número decimal pertencente ao intervalo [-1,1], onde o 1 significa que os dois *patches* são idênticos. Já a função de perda com SSIM usada para treinamento da rede é dada por:

$$l_{SSIM} = 1 - SSIM(w, z). \tag{4.3}$$

No início do treinamento, espera-se que as imagens aprimoradas pela rede apresentem diferenças significativas em relação ao *ground truth*, fazendo com que a medida SSIM retorne valores próximos a zero. Por isso, a função de perda l_{SSIM} subtrai esse resultado do valor 1, para que o erro (informação usada para corrigir os pesos da rede) seja maior nesses casos.

Uma das formas para analisar, de forma automática, a qualidade obtida nos processos de aprimoramento em sinais é através do método de referência completa, o qual requer o sinal original não degradado para fazer uma comparação com o sinal resultante (MELLO et al., 2017).

Quanto à parte de avaliação dos resultados obtidos pela rede, foram utilizadas as medidas SSIM, PSNR (sigla do inglês *Peak Signal-to-Noise Ratio*) e LOE (sigla do inglês *Lightness Order Error*) (WANG et al., 2013).

O PSNR é usado como medida de qualidade entre a imagem original e a reconstruída. Quanto maior o valor retornado por essa medida, melhor a qualidade da segunda imagem. A relação calculada pelo PSNR sobre duas imagens E e E' pode ser definida como:

$$PSNR = 10 \cdot log_{10}\Big(\frac{MAX_E^2}{MSE}\Big), \tag{4.4}$$

sendo MAX_E o valor máximo possível para pixel segundo o tipo de dados da imagem de entrada e MSE correspondente ao erro médio quadrático (MSE do inglês: *Mean Squared Error*) dado por:

$$MSE = \frac{\sum_{M,N}[E(m,n) - E'(m,n)]^2}{M * N}, \tag{4.5}$$

onde M e N são as quantidades de linhas e colunas da imagem de entrada.

Segundo Wang et al. (2013) a naturalidade de uma imagem aprimorada está relacionada à direção e variação da luz ao longo da imagem. Sendo estas informações representadas pela ordem relativa da luminosidade, Wang et al. (2013) propõem a medida LOE (sigla do inglês *Lightness Order Error*) que é definida com base no erro dessa ordem entre a imagem original E e sua versão aprimorada E'.

A luminosidade $I(p)$ de uma imagem no pixel p é dada como o valor máximo de seus três canais de cor:

$$I(p) = \max_{c \in \{r,g,b\}} E^c(p) \tag{4.6}$$

Para cada pixel p, a diferença de ordem relativa à luminosidade entre E e E' é dada por:

$$RD(p) = \sum_{k=1}^{t} U\Big(I(p), I(k)\Big) \oplus U\Big(I'(p), I'(k)\Big), \tag{4.7}$$

onde t representa a quantidade de pixels, I(p) e I'(p) indicam a luminosidade no pixel p na imagem original e aprimorada, respetivamente. Já o operador $\oplus$ representa a operação de ou exclusivo e a função $U(a, b)$ retorna 1 se $a \geq b$ e 0, caso contrário. Por fim, a medida LOE é definida como:

$$LOE = \frac{1}{t} \sum_{p=1}^{t} RD(p). \tag{4.8}$$

5 EXPERIMENTOS

5.1 EXPERIMENTOS EM CONSTÂNCIA DE COR

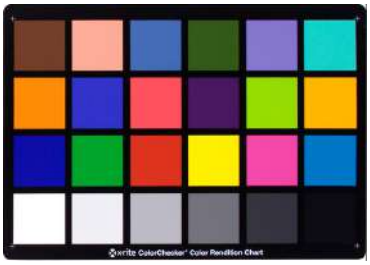
As modificações propostas à rede FC4 foram implementadas sobre o código disponibilizado pelos próprios autores [1] e com auxílio da biblioteca *TensorFlow* (ABADI et al., 2015). Os programas foram executados em uma máquina com placa gráfica NVIDIA TITAN XP e para garantir o determinismo sobre os testes foi utilizada a biblioteca *TensorFlow Determinism* v0.3 [2].

A seguir, são descritos os principais experimentos realizados. Os primeiros testes focaram na escolha do mapa de saliência mais adequado ao problema. Com o tipo de mapa já definido, foram executados experimentos a fim de verificar em quais aspectos a inclusão do mapa na rede FC4 poderia trazer vantagens ao sistema.

5.1.1 Base de Dados

O conjunto de imagens utilizado nos experimentos foi o da base de Gehler (SHI, 2000), o qual contém 568 imagens de alta qualidade capturadas por duas câmeras DSLR (sigla do inglês *Digital Single-Lens Reflex*) em ambientes tanto internos quanto externos. Em cada imagem da base, está presente uma placa (ver Figura 40) chamada de *Macbeth ColorChecker* (MCC) que é usada para fornecimento dos valores verdadeiros (*ground truth*) das cores da iluminação. O *colorchecker* possui seis quadrados acromáticos e é com base neles que ocorre o cálculo do *ground truth* da iluminação. Esse cálculo envolve as medianas de cada canal RGB do quadrado acromático mais brilhante.

Figura 40 – *Colorchecker*



Fonte: Site da Amazon[3].

As coordenadas do *colorchecker*, dadas pelo próprio *dataset*, são utilizadas para mascarar a região referente ao mesmo com pixels pretos (ver Fig. 41). A aplicação dessa máscara

[1] Disponível em: <https://github.com/yuanming-hu/fc4/>
[2] Disponível em: <https://github.com/NVIDIA/tensorflow-determinism>
[3] Disponível em: <https:https://www.amazon.co.uk/X-Rite-ColorChecker-Classic-color-Rendition/dp/B000JLO31C>.

é feita para evitar que o treinamento e o teste da rede sejam influenciados, de alguma forma, por esse objeto. Assim como em trabalhos anteriores, a base é avaliada utilizando validação cruzada com três *folders*.

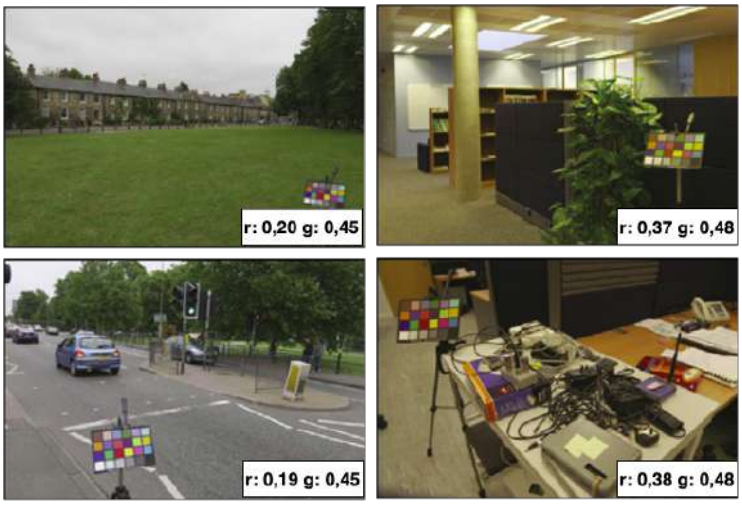
Figura 41 – Aplicação de máscara sobre região do *colorchecker*.

(a) Imagem de entrada (b) imagem com máscara

Fonte: O autor.

A Figura 42 apresenta algumas das imagens pertencentes ao conjunto; nela percebe-se que cenas capturadas sob condições parecidas tendem a apresentar iluminações mais próximas. Por exemplo, ambas imagens na Fig. 42a são de ambientes externos em um dia nublado, enquanto que as da Fig. 42b remetem a locais fechados iluminados aparentemente por tipos semelhantes de lâmpada. Em cada imagem, são expostos os valores dos tons r e g (vermelho e verde) da iluminação.

Figura 42 – Imagens da base Gehler com iluminações semelhantes.



(a) (b)

Fonte: Adaptado de (OH; KIM, 2017).

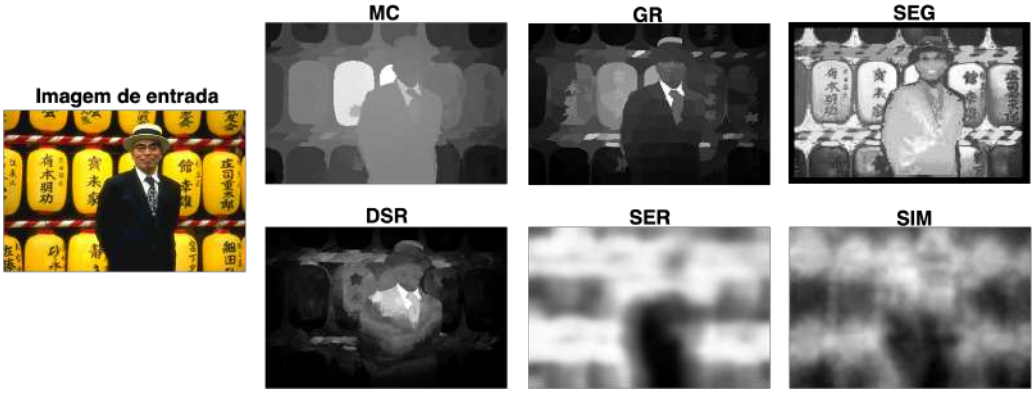
5.1.2 Escolha do mapa de saliência

Para esses testes, foi utilizada a revisão feita por Borji et al. (2014) sobre modelos para detecção de regiões salientes como um ponto de partida para escolha do mapa, uma vez que

a quantidade de algoritmos com esse objetivo é consideravelmente grande na literatura.

Os experimentos consistiram em submeter, para cada método de detecção de saliência, cerca de 20 imagens da base BSDS500 (MARTIN et al., 2001) (ver Apêndice A com alguns exemplos) escolhidas de forma a tentar abranger casos o mais diferentes possíveis.A Figura 43 exemplifica os mapas obtidos pelas técnicas MC (JIANG et al., 2013), GR (YANG; ZHANG; LU, 2013), SEG (RAHTU et al., 2010), DSR (LI et al., 2013), SeR (SEO; MILANFAR, 2009) e SIM (MURRAY et al., 2011) para uma mesma imagem de entrada. De forma geral, percebeu-se que a forma como cada método detecta regiões salientes visualmente segue um padrão. Por exemplo, o método SeR tende a não ser preciso na delimitação das regiões importantes, enquanto que a técnica SEG o faz com muita precisão. Sobre esses resultados, foi feita uma análise visual e subjetiva, na qual os métodos MC e GR foram considerados, de maneira geral, os que proporcionaram detecções mais próximas daquelas que se esperava. Uma vez que as regiões detectadas como salientes não são representadas de forma muito abstrata (como fazem as técnicas SeR e SIM, por exemplo), nem com muitos detalhes como a abordagem SEG.

Figura 43 – Comparação entre mapas de saliência.



Fonte: O autor.

Normalmente, os algoritmos baseados em CNN para correção de cor mascaram com o valor 0 as regiões correspondentes aos objetos utilizados para calibração da cor. No caso da base de dados utilizada, o objeto a ser mascarado é o *colorchecker*, como mencionado anteriormente. No entanto, percebeu-se que, comumente, a região referente à placa nas imagens é detectada por ambos os métodos de saliência como de destaque. Desta forma, aplicou-se um ajuste a fim de garantir-se o mascaramento do *colorchecker* também nos mapas. Esse ajuste pode ser feito, utilizando as coordenadas referentes aos quatro cantos da placa (disponibilizadas pela própria base) para criação de um polígono responsável por ocultar qualquer possível importância dada inicialmente a esse local. Um exemplo de aplicação desse ajuste é apresentado na Figura 44.

Os mapas de saliência, depois de ajustados, foram submetidos juntamente com as imagens da base Gehler para treinamento da rede com duração de 50 épocas. A Figura

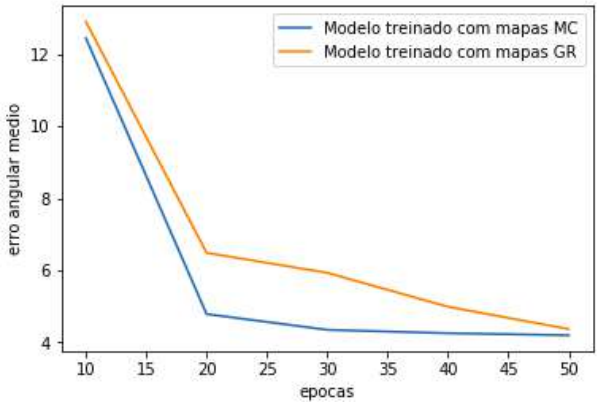
Figura 44 – Exemplo de ajuste feito sobre saída de método MC.

(a) imagem de entrada (b) mapa de saliência inicial (c) mapa ajustado

Fonte: O autor.

46 traz o gráfico do desempenho da rede quanto à média do erro angular calculado sobre as estimativas de iluminação. Nele, percebe-se que os modelos apresentam curvas de erro muito próximas, permitindo, a princípio, a escolha de qualquer um dos dois. Porém, ao aumentar-se o tempo de treinamento para 100 épocas, o modelo com mapas MC obteve menor erro angular médio sendo, portanto, o escolhido para a próxima fase de experimentos que é discuta na próxima seção.

Figura 45 – Comparação entre mapas MC e GR.



5.1.3 Testes e análise de resultados

Seguindo trabalhos anteriores, várias medidas padrão são aplicadas em relação ao erro angular, sendo elas: média, mediana, média dos 25% dos casos de maior acerto e 25% dos de maior erro. Aplicando-se a validação cruzada com três *folders* sobre as 568 imagens da base Gehler, os modelos foram treinados durante 1000 épocas, organizados de forma decrescente quanto ao erro. Como já observado anteriormente, a versão original da FC4 que separa o mapas de confiança dos de estimativa quanto à iluminação (FC4 V1) foi a que apresentou maior dificuldade em aprender, retornando uma diferença de mais $1,1°$ em relação ao modelo modificado que, por aplicar a mesma separação entre confiança e estimativa, lhe é mais semelhante: mFC4 V1. Um dos motivos que pode estar favorecendo

essa diferença é o comportamento de uma função, chamada *softmax*, sobre o mapeamento de pesos. A *softmax* (idealmente usada nas últimas camadas de redes para classificação) é responsável por dado um vetor de números reais normalizá-los entre 0 e 1. Como a soma dos novos valores normalizados é sempre igual a 1, a *softmax* é considerada como uma função de distribuição de probabilidades. Voltando ao caso da FC4 V1, percebeu-se que esta função quando aplicada aos mapas de confiança (já linearizados) apresentou uma tendência em acumular maior probabilidade apenas para aproximadamente dois ou três *patches*. Isso, na prática, significa que boa parte da imagem terá pouca ou nenhuma representatividade para o cálculo do iluminante global, o que ajuda a entender o porquê do erro angular ser alto para o modelo FC4 V1.

Tabela 2 – Resultados de modelos sobre base Gehler.

Método	**Média**	**Mediana**	**25% melhores**	**25% piores**
FC4 V1	3,690	2,846	0,957	8,0717
mFC4 V2	2,950	2,126	0,897	6,377
mFC4 V1	2,582	2,000	0,773	5,301
FC4 V2	2,109	1,638	0,605	4,429

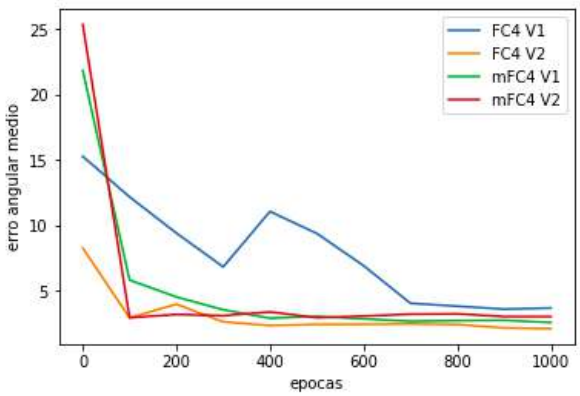
A modificação sugerida mFC4 V1 consiste na completa substituição do mapa de confiança $\widehat{c}$ (gerado pela rede) pelo mapa de saliência m. Um dos objetivos com essa substituição era analisar o potencial de algoritmos de detecção de regiões salientes como mecanismo para identificação de *patches* informativos e o quanto isso impactaria no cálculo do iluminante global. Testes mostram que o uso desse tipo de informação é útil na discriminação dos *patches* que mais podem colaborar para o balanceamento de cores, superando uma das versões *baseline* (FC4 v1) com uma diferença de mais de 1° para o erro angular médio. No entanto, quando comparado com a versão que assume que a confiança $\widehat{c}_i$ pode ser dada pela norma do vetor de iluminação $\widehat{p}_i$ (FC4 V2), a rede modificada perde com uma diferença de $0,473°$.

Por sua vez, a rede mFC4 V2 propõe a multiplicação do mapa de confiança $\widehat{c}$ pelo de saliência m como agente ponderador para as estimativas locais de iluminação $\widehat{p}_i$. Os resultados apresentados pela Tabela 2 indicam que essa combinação trouxe benefícios para rede, diminuindo em $0,74°$ a média do erro angular em relação à FC4 V1 que usa apenas a confiança $\widehat{c}_i$. Como não houve modificação direta quanto à geração do mapa de iluminação $\widehat{p}$, supõe-se que essa queda é devido a uma melhor distribuição de pesos entre as regiões da imagem. Todavia, o erro médio da mFC4 V2 ainda é superior ao da versão

original FC4 V2.

A Figura 46 traz o comportamento da curva de erro para cada modelo ao longo das 1000 épocas, permitindo assim uma certa visualização, embora limitada, de como foi o processo de aprendizagem para cada rede. Frente a essas curvas de erro dos modelos, percebeu-se uma significativa diferença no desempenho das duas versões da FC4, que pode ser observada pela distância entre suas curvas. Embora ambas tenham sido colocadas pelos autores como equivalentes, a versão 1 precisaria de, no mínimo, maior tempo de treinamento para chegar no nível de precisão da segunda. Além disso, observou-se que as modificações propostas não conseguiram diminuir a taxa de erro obtida pela rede original em sua melhor configuração, chegando apenas de forma muito próxima.

Figura 46 – Curvas de erro sobre a base Gehler.



.

Os valores apresentados pela Tabela 2 são referentes às redes treinadas localmente com um total de 1000 épocas. Contudo esses resultados ainda se mostram distantes se comparados aos dos métodos mais recentes (ver Tabela 3). Por isso, optou-se por aumentar o tempo de treinamento da variação mFC4 V1 que, pelos testes anteriores, apresentou melhor desempenho, de 1000 para 6000 épocas ao longo das quais o erro angular médio caiu de 2,582 para 2,083, continuando a ser superior ao erro reportado no trabalho de Hu, Wang e Lin (2017).

Tais resultados indicam que, pelo menos na base Gehler, a inclusão de mapas de saliência como mecanismo para ponderação de *patches* pode não ser a melhor forma de identificar regiões mais significativas para o cálculo do iluminante.

Na próxima seção são apresentadas algumas situações em que a variação proposta mFC4 V1 obteve melhores resultados que a *baseline* FC4 V2, indicando que talvez o uso da primeira possa ser futuramente combinado com a última em seus casos de falha.

Tabela 3 – Estatísticas sobre o erro angular obtidas por alguns dos métodos pertencentes ao estado da arte sobre o conjunto Gehler.

Método	**Média**	**Mediana**	**25% melhores**	**25% piores**
White Patch Retinex	7,55	5,68	1,45	16,12
Gray World	6,52	5,04	1,90	13,58
(BIANCO; CUSANO; SCHETTINI, 2017)	2,36	1,44	-	-
(OH; KIM, 2017)	2,16	1,47	0,37	5,12
mFC4 V1	**2,08**	**1,54**	**0,49**	**4,63**
FC4	1,65	1,12	0,38	3,78

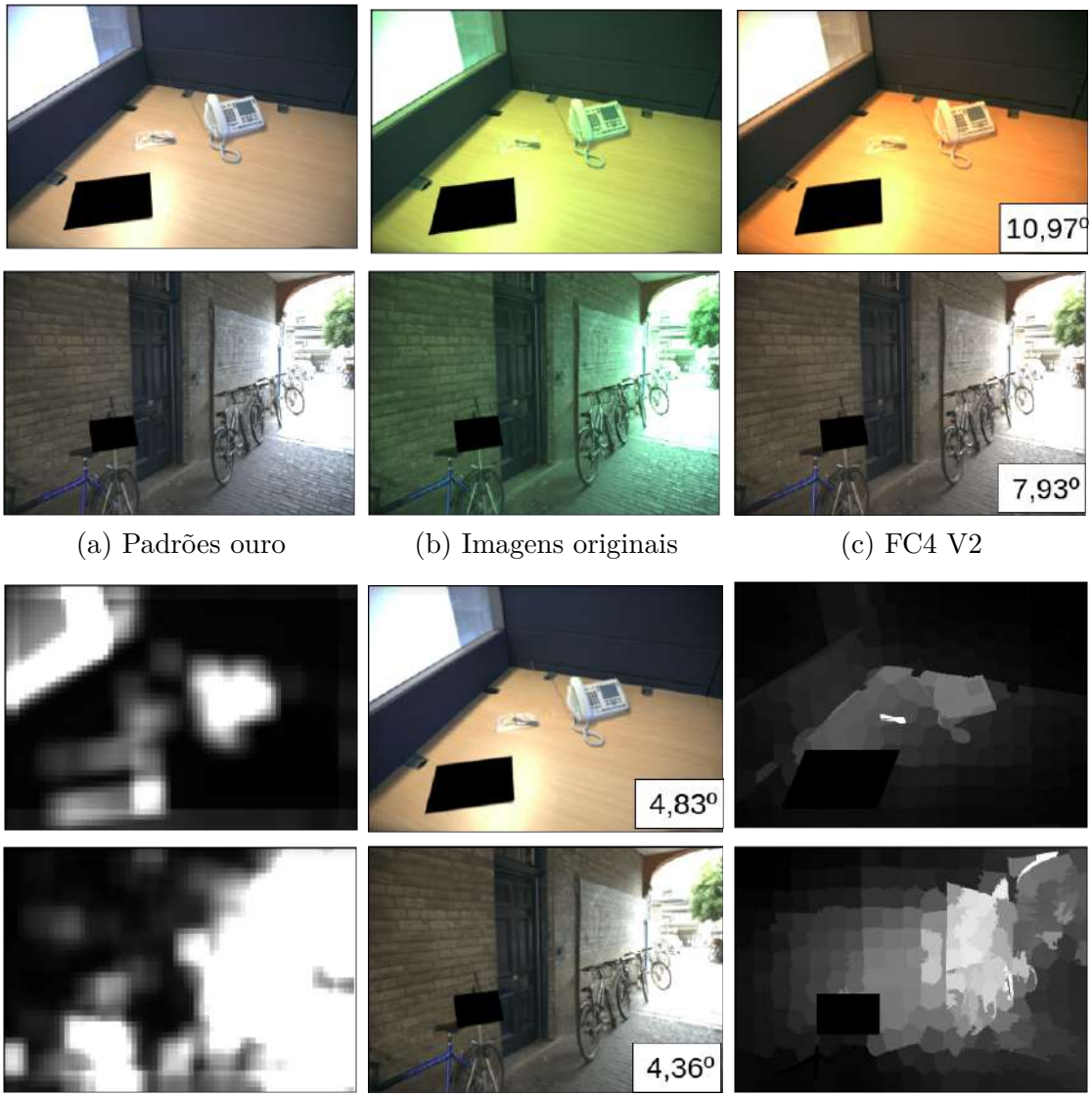
5.1.4 Estudo de casos

Hu, Wang e Lin (2017) apresentam alguns elementos que geralmente recebem uma alta confiança, ou seja, são considerados pela rede como bons informantes sobre a iluminação. São eles: reflexos especulares, superfícies com textura rica, objetos com cores inatas, faces humanas e manchas brilhantes. Após uma análise dos mapas de confiança e de saliência de alguns dos casos que envolvem os dois últimos aspectos (faces e regiões brilhantes), algumas observações foram feitas e sobre estas discute-se a seguir.

Quanto à presença de regiões de intenso brilho, a Figura 47 traz alguns dos resultados onde a FC4 apresentou maior erro, ainda que os mapas de confiança tenham atribuído maior peso para *patches* mais brilhantes. Nesses casos, percebeu-se que os mapas de saliência (Figura 47f), menos influenciados por tal condição, conseguiram proporcionar erros angulares inferiores e, consequentemente, imagens com cores mais próximas às indicadas pelo *ground truth* (Fig. 47a). Supõe-se que um dos motivos para essas ocorrências seja a existência de possíveis restrições quanto à intensidade e/ou área abrangida por intensa luminosidade, de forma que quando há um "excesso"em um desses aspectos a eficácia da distribuição dos pesos, representada pelo mapa de confiança $\widehat{c}$, fica comprometida. Todavia, não foi ainda possível confirmar essa suspeita por meio de testes quantitativos para toda a base.

Outra característica analisada foi a presença de faces humanas e sua contribuição para o cálculo do iluminante. A Figura 48 traz alguns casos onde locais com rostos localizados em regiões de sombra ou penumbra não foram bem ponderadas pela rede FC4. Esse comportamento, de certa forma, é entendível se a confiança por *patch* for dada pela norma do vetor de iluminação (caso da FC4 V2). De forma que quanto mais clara for a

Figura 47 – Casos de imagens com regiões de intenso brilho.



(a) Padrões ouro (b) Imagens originais (c) FC4 V2

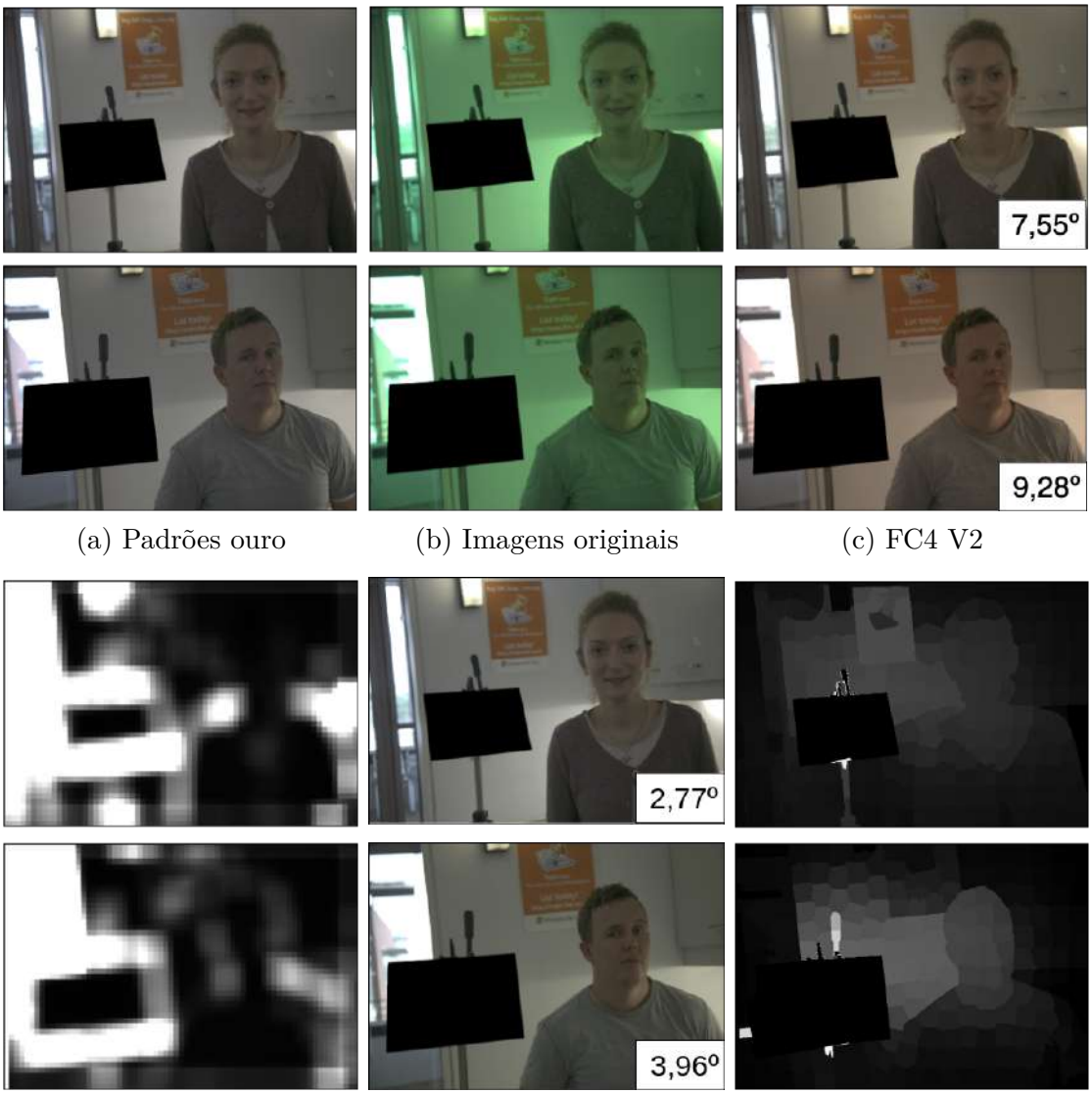
(d) Mapas de confiança (e) mFC4 V1 (f) Mapas de saliência

Fonte: O autor

região, maior será o peso atribuído. Todavia, isso pode limitar o aproveitamento de objetos da cena informativos quanto a cor, que por estarem em locais não tão bem iluminados, acabam sendo ignorados pela rede.

Por esta razão, no próximo capítulo, são apresentados alguns métodos voltados ao aprimoramento de imagens com baixa luminosidade. A ideia é futuramente utilizar esse tipo de correção como uma etapa de pré-processamento em sistemas para constância de cor, para que redes como a FC4 (influenciadas por regiões de maior brilho) não deixem de se beneficiar por *patches* informativos ainda que estes estejam sombreados.

Figura 48 – Casos de imagens com face humana.



(a) Padrões ouro (b) Imagens originais (c) FC4 V2

(d) Mapas de confiança (e) mFC4 V1 (f) Mapas de saliência

Fonte: O autor

5.2 EXPERIMENTOS EM IMAGENS FRACAMENTE ILUMINADAS

Um dos maiores desafios encontrados para o treinamento de redes profundas no tipo de problema em questão é justamente a falta de dados rotulados, ou seja, pares de imagens compostos pela versões degradada e ideal de iluminação. Para contornar esse problema, as imagens escurecidas são sintetizadas pela Equação 4.1; ou seja, dada uma imagem clara e bem iluminada $R(x, y)$ e uma iluminação aleatória I, a imagem fracamente iluminada $E(x, y)$ é gerada multiplicando-se $R(x, y)$ por I. A Figura 49 traz alguns exemplos desse processo de síntese.

Foram selecionadas 470 imagens claras das 500 disponíveis na base BSDS500 (ARBELAEZ et al., 2010) para atuarem como *ground truth*. Isto porque, trinta imagens presentes

na base foram capturadas sob condições de baixa iluminação (e.g. ambientes noturnos e regiões do fundo do mar). A iluminação I é, então, tomada de forma aleatória a partir do intervalo $[0,;0,5]$ (quanto menor for o valor para I, maior será o efeito de escurecimento). As dimensões dos *patches* gerados foram mantidas em 40×40, já que em (ZHANG et al., 2017) há relato da relação desse parâmetro com a profundidade da rede. Para prevenir o *overfitting*, foram aplicadas também algumas transformações sobre as imagens de treinamento como rotação, escala e espelhamento.

Figura 49 – Exemplos de imagens bem iluminadas (primeira linha) e suas versões fracamente iluminadas (segunda linha).

O desempenho da rede WIIEN é avaliado sobre dois conjuntos de dados. O primeiro consiste em 24 imagens provenientes da base Kodak (FRANZEN, 1999), as quais foram degradadas com base na Equação 4.1 com a iluminação sendo escolhida de forma aleatória dentro do intervalo [0,; 0,5]. Já o segundo conjunto é composto por imagens reais de baixa luminosidade, sendo elas de cinco *datasets* públicos: VV (VONIKAKIS; KOUSKOURIDAS; GASTERATOS, 2018), LIME (GUO; LI; LING, 2016), NPE (WANG et al., 2013), DICM (LEE; LEE; KIM, 2012) e MEF (MA; ZENG; WANG, 2015). Além disso, os resultados obtidos pela rede WIIEN sobre tais bases são comparados com alguns métodos do estado da arte: NPE (WANG et al., 2013), Dong (DONG et al., 2011), MEF (MA; ZENG; WANG, 2015), LIME(GUO; LI; LING, 2016), SRIE (FU et al., 2016) e BIMEF (YING; LI; GAO, 2017). A seguir são relatados os resultados obtidos para cada conjunto de teste.

5.2.1 Testes em base sintética

Por ser difícil encontrar imagens de baixa luminosidade com o *ground truth* disponível, foram usadas imagens sintéticas tanto para o treinamento, como para parte dos testes. A base sintética é composta por 24 cenas, em sua maioria, capturadas em boas condições

de iluminação. Seguindo o trabalho de (LI et al., 2018), aplicou-se um valor aleatório para a iluminação I a fim de simular ambientes com baixa intensidade desse componente.

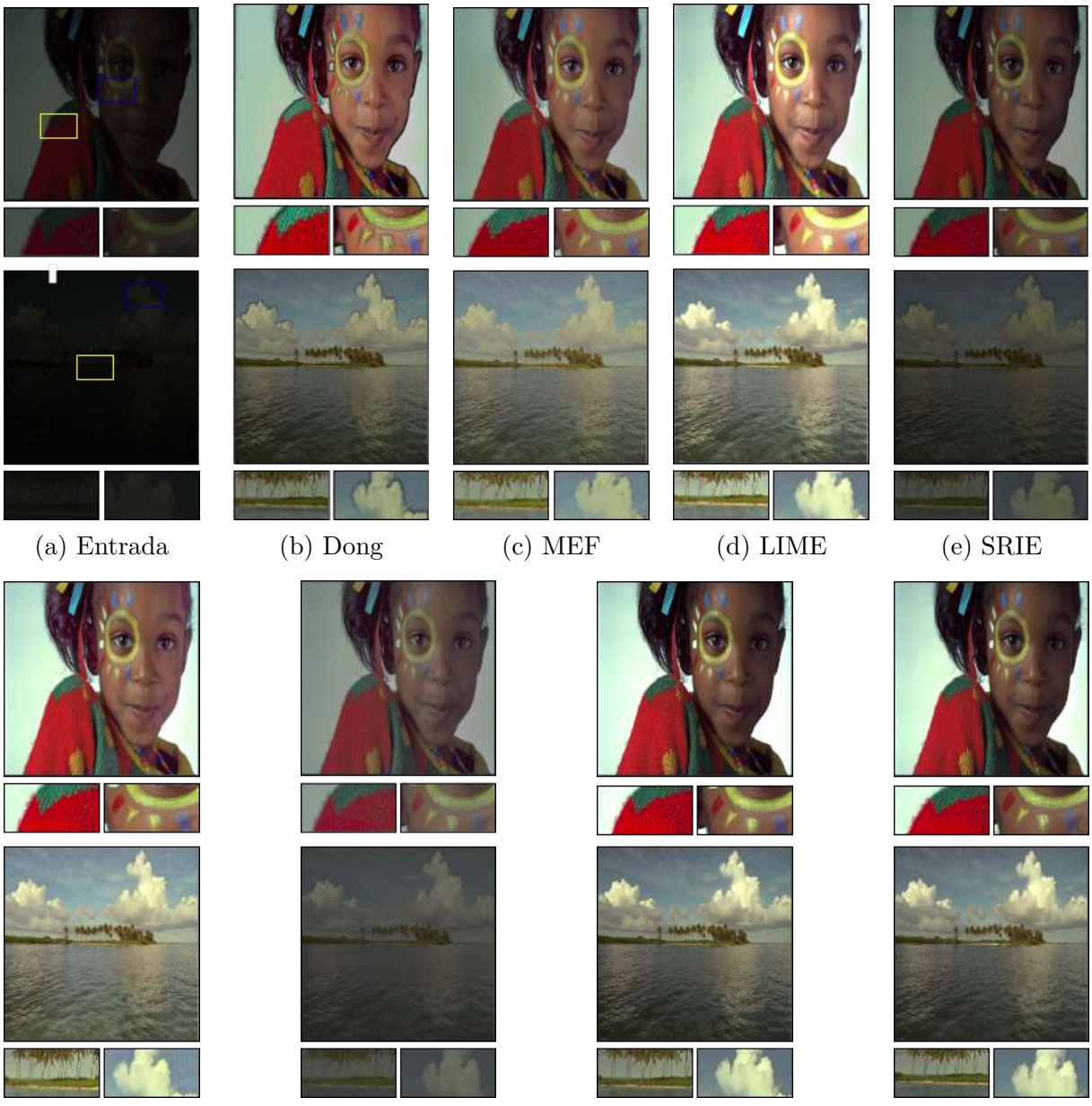
O conjunto de medidas usado para avaliação de qualidade das imagens aprimoradas pela rede WIIEN incluiu os índices PSNR, SSIM e LOE. A Tabela 4 mostra o desempenho, em termos quantitativos, para base sintética. O primeiro, segundo e terceiro melhores casos, para cada medida, são destacados de vermelho, verde e azul, respectivamente. Percebe-se que o trabalho proposto recebeu melhor avaliação em todas as medidas, além de apresentar uma diferença significativa em relação aos métodos do estado da arte.

Alguns resultados representativos são ilustrados pela Figura 50. Em particular, percebe-se que o método Dong gera traços irreais na cena, como o contorno preto ao longo da tinta e da nuvem. Por sua vez, o LIME super realçou as cores (ver intensidade da tinta amarela) proporcionando até um resultado visualmente agradável, mas que está longe do original. Enquanto isso, os métodos MEF, SRIE e BIMEF foram os que apresentaram menor nível de contraste e brilho em relação aos demais. Por fim, como já indicado pela Tabela 4, a abordagem NPE mostrou bom desempenho em termos de luminância e contraste. É possível ver que a questão do sombreamento no rosto da menina (Fig. 50f) foi amenizado e esse aspecto pode ser estratégico dependendo da aplicação. No entanto, todos esses métodos do estado da arte não conseguiram ser mais fiéis do que a rede WIIEN em relação às cores das imagens originais. Isso pode ser observado, por exemplo, quando verifica-se a distorção na cor de pele por cada método.

Tabela 4 – Comparação entre diferentes métodos sobre base sintética. O primeiro, segundo e terceiro melhores casos, para cada medida, são destacados de vermelho, verde e azul, respectivamente.

Método	**PSNR**	**SSIM**	**LOE**
Dong	18,27	0,8320	1592
NPE	21,01	0,9177	588,29
LIME	16,96	0,8562	854,13
MEF	20,80	0,9182	910,66
SRIE	17,69	0,8701	751,18
BIMEF	18,43	0,8641	348,00
WIIEN	**25,97**	**0,9720**	**174,83**

Figura 50 – Exemplos de imagens sintéticas com baixa luminosidade e suas versões aprimoradas.



(a) Entrada (b) Dong (c) MEF (d) LIME (e) SRIE

(f) NPE (g) BIMEF (h) WIIEN (i) original

5.2.2 Testes em bases reais

Além do conjunto de imagens sintéticas, também foram submetidas à rede WIIEN bases reais de baixa luminosidade as quais por não disponibilizarem *ground truth* quanto à iluminação inviabilizam o uso das medidas PSNR e SSIM, sendo aplicada somente a medida LOE.

Conforme apresentado na Tabela 5, o algoritmo proposto obtém melhores resultados em todas as bases reais quando comparado aos do estado da arte. Isto significa que a rede WIIEN tem capacidade para fazer o aprimoramento quanto à luminosidade de uma imagem mantendo, ao mesmo tempo, a naturalidade da mesma.

Conforme apresentado na Tabela 5, o algoritmo proposto obtém melhores resultados em todas as bases reais quando comparado aos do estado da arte. Assim como nos resultados para base sintética, o primeiro, segundo e terceiro melhores casos são destacados de vermelho, verde e azul, respectivamente. Para esse conjunto de imagens, a rede proposta também manteve os melhores resultados, embora que com uma diferença menor em relação ao segundo melhor (BIMEF).

Alguns exemplos de imagens reais fracamente iluminadas e seus resultados gerados por cada método são ilustrados na Fig. 51. As cenas aprimoradas por Dong continuam a apresentar contornos super realçados. Além disso, foram percebidas distorções de cores nos resultados das abordagens Dong, LIME, e NPE (um trecho do asfalto adquiriu manchas rosas). Todavia, também percebeu-se que embora o erro obtido pela rede WIIEN tenha sido o menor, a mesma não conseguiu clarear certos locais da imagem que os demais métodos conseguiram (mesmo que estes, em sua maioria, acabaram adicionando ruído em tais regiões). Casos de falha como esses já eram esperados uma vez que o iluminante estimado pela rede é global e não local, sendo mais prováveis em imagens onde há um forte contraste de iluminação.

Tabela 5 – Resultados quanto à medida de distorção da luz (LOE). O primeiro, segundo e terceiro melhores casos, para cada medida, são destacados de vermelho, verde e azul, respectivamente.

Método \ **Base**	**LIME**	**NPE**	**MEF**	**VV**	**DICM**
Dong	1192,3	955,745	936,52	1057,8	1171,9
NPE	1510,7	647,05	1131.6	916,37	663,78
LIME	1330,8	1141,1	1159,8	11179	12636
MEF	553,63	481,39	469,62	471,74	618
SRIE	811,07	553,98	755,67	721,71	626.40
BIMEF	365,55	254,21	285,64	280,31	353,27
WIIEN	**141,50**	**128,58**	**154,7**	**211,24**	**247,54**

Figura 51 – Exemplos de imagens reais com baixa luminosidade e suas versões aprimoradas.

(a) Entrada

(b) Dong

(c) MEF

(d) LIME

(e) SRIE

(f) NPE

(g) BIMEF

(h) WIIEN

6 CONCLUSÃO

Nesta dissertação, abordam-se dois tipos de aprimoramento de imagens, sendo o primeiro voltado para o problema de constância de cor e o segundo, para aprimoramento de imagens de baixa luminosidade. Durante o desenvolvimento de tais métodos, os objetivos específicos foram atingidos conforme relata-se a seguir.

Para a questão da inconsistência de cores, alguns dos principais métodos de estimativa de iluminante clássicos (*white patch* e *gray world*, por exemplo) e modernos (técnicas baseadas em aprendizagem profunda) foram estudados e testados. De todas as abordagens analisadas, destacou-se a rede FC4 proposta por Hu, Wang e Lin (2019), a qual traz uma nova estratégia de combinação entre estimativas locais de iluminação para geração de uma estimativa global. Partindo da suposição que existem regiões da imagem que fornecem mais informações quanto ao iluminante que outras, o método original propõe uma rede capaz de calcular o peso que cada estimativa local tem para o cálculo da iluminação global. No entanto, após analisar esse ponderamento em diversas imagens da base Gehler (SHI, 2000), percebeu-se que a rede acaba tendo uma tendência em concentrar maior peso em regiões super brilhantes da imagem, como, por exemplo, janelas ou portas com forte claridade solar, desprezando muitas vezes boa parte do restante da cena.

Dado esse cenário, neste trabalho foram propostas e desenvolvidas duas modificações à rede de Hu, Wang e Lin (2019) com o intuito de melhorar o desempenho da método original. Ambas variações envolvem a inclusão de mapas de saliência, seja atribuindo a eles a completa responsabilidade pela distribuição de peso entre os *patches* (mFC4 V1) ou colocando-os de forma colaborativa com o mapa de confiança gerado pela rede FC4 (mFC4 V2). A escolha por mapas de saliência gerados pelo método MC (JIANG et al., 2013) se deu por estes demonstrarem uma melhor distribuição de confiança entre os *patches*, mesmo em presença de locais de intenso brilho.

Os experimentos foram executados aplicando-se as duas modificações propostas, bem como as duas versões da rede FC4 sobre a base Gehler, tendo como objetivo principal verificar se a inclusão dos mapas de saliência, de fato, beneficiaria o cálculo da iluminação. Seguindo trabalhos anteriores, a medida utilizada nos testes foi a do erro angular, por esta não depender da escala utilizada para representação da iluminação, facilitando, assim, a comparação entre os algoritmos para constância de cor.

Os resultados obtidos mostraram que a inclusão de mapas de saliência na rede FC4 não proporcionou menor erro médio que a versão original, na qual os pesos são implicitamente representados pelas normas dos iluminantes locais (FC4 V2). No entanto, é válido ressaltar que até mesmo para a rede FC4 V2 a maior diferença de erro em relação a versão modificada não chegou a 1°, no caso do treinamento de 1000 épocas. Por sua vez, o estudo de casos apontou algumas situações, e.g. locais de brilho intenso e/ou rosto humano, em

que a rede mFC4 V1 apresentou melhores resultados tanto em termos de erro angular, como em cores mais próximas ao que seria ideal.

A análise do estudo de caso também permitiu identificar que *patches* contendo rosto humano (considerados pela rede como bons informantes sobre a iluminação) receberam menor confiança por estarem em regiões não tão bem iluminadas. Nesse contexto, buscou-se entender mais sobre como funcionam os métodos de aprimoramento de imagens com baixa luminosidade. Na perspectiva de que os mesmos possam ser aplicados futuramente como pré-processamento, fazendo com que objetos informativos presentes em regiões escuras tenham maior visibilidade para redes como a FC4 (influenciadas por locais de maior brilho).

O outro tipo de aprimoramento proposto neste trabalho é voltado para imagens fracamente iluminadas. Normalmente, imagens dessa natureza apresentam histogramas concentrados em níveis baixos de intensidade, implicando em pouco contraste na cena. Tais técnicas tendem a melhorar a identificação de objetos ou texturas localizadas em regiões não tão bem iluminadas, podendo ser estratégicas em aplicações de visão computacional.

Visto isso, foi desenvolvida uma metodologia para o aprimoramento de imagens fracamente iluminadas. O método proposto, chamado de WIIEN (sigla do inglês: *Weakly Illuminated Image Enhancement Network*), é baseado em uma rede CNN cujo objetivo consiste primeiramente em estimar o mapa de iluminação a partir da imagem de entrada, calculando a partir desse mapa um valor global para iluminação, para, só então, utilizá-lo na correção da mesma. Seguindo trabalhos anteriores como o de Li et al. (2018), a rede WIIEN adota valores escalares para iluminação. Dessa forma, diminui-se a probabilidade do método em gerar distorções de cores na cena.

Os experimentos contaram com um conjunto de imagens sintéticas e outro composto por cenas reais de baixa luminosidade. Quanto à validação do método, foram aplicadas as medidas SSIM, PSNR e LOE sobre as imagens aprimoradas. Tais medidas, comumente encontradas em trabalhos do estado da arte, quantificam a qualidade da imagem resultante em termos de estrutura e naturalidade da cena. Sobre ambos os conjuntos de dados, a rede WIIEN recebeu os melhores resultados em todas as medidas quando comparada aos métodos do estado da arte. Além disso, também observou-se a tendência da rede proposta em preservar as cores da imagem.

No entanto, a abordagem proposta apresenta algumas limitações. A principal delas consiste na utilização de um iluminante global para correção, o que acaba comprometendo a eficiência da mesma em melhorar a visibilidade de cenas com regiões de extremos brilho e escuridão.

6.1 CONTRIBUIÇÕES

A primeira contribuição deste trabalho está na investigação sobre o uso de mapas de saliências como mecanismo discriminativo de regiões da imagem consideradas mais in-

formativas quanto à cor da iluminação. Embora que os resultados não tenham superado aos da metodologia tomada por base, o uso desse tipo de mapeamento não prejudicou o sistema a ponto de fazê-lo mais falho que outros do estado da arte. Não distante disso, observaram-se algumas situações específicas onde sua aplicação trouxe vantagens para a rede.

Já a segunda contribuição consiste na apresentação de uma abordagem baseada em uma rede neural convolucional para o aprimoramento de imagens fracamente iluminadas. Cenas reais com diversas variações de iluminação foram submetidas à rede que, por sua vez, proporcionou cenas com maior visibilidade sem com isso comprometer a naturalidade da mesma.

Nessa aplicação, uma base de imagens sintéticas foi desenvolvida e será disponibilizada na versão final desta Dissertação. Entendemos esta como uma terceira contribuição deste trabalho.

6.2 TRABALHOS FUTUROS

A partir desta dissertação, alguns trabalhos futuros podem ser propostos como:

1. Incluir como pré-processamento para rede FC4 técnicas de aprimoramento para imagens de baixa luminosidade como, por exemplo, a rede WIIEN. A ideia seria verificar se o aperfeiçoamento na visibilidade da cena pode contribuir para redução do erro quanto à estimativa de iluminação;

2. Modificar rede WIIEN para que esta possa corrigir as imagens com base em estimativas locais de iluminante. Dessa forma, casos onde a iluminação não é uniforme poderiam ser melhor tratados;

3. Aplicar medidas voltadas para análise de contraste nos resultados obtidos pela abordagem WIIEN;

4. Adaptar o método WIIEN para o problema de constância de cor. Ao invés da estimar-se o iluminante como um escalar, considerá-lo como um vetor onde cada valor seria referente a um canal de cor.

5. Testar novos modelos de mapas de saliência.

6. Analisar os resultados alcançados para detectar padrões, onde os métodos de mapas de saliência conseguem alcançar melhores resultados do que os mapas de confiança. Dessa forma, pode-se, previamente, definir a melhor forma de otimizar os pesos por um mapa ou por outro.

REFERÊNCIAS

ABADI, M.; AGARWAL, A.; BARHAM, P.; BREVDO, E.; CHEN, Z.; CITRO, C.; CORRADO, G. S.; DAVIS, A.; DEAN, J.; DEVIN, M.; GHEMAWAT, S.; GOODFELLOW, I.; HARP, A.; IRVING, G.; ISARD, M.; JIA, Y.; JOZEFOWICZ, R.; KAISER, L.; KUDLUR, M.; LEVENBERG, J.; MANé, D.; MONGA, R.; MOORE, S.; MURRAY, D.; OLAH, C.; SCHUSTER, M.; SHLENS, J.; STEINER, B.; SUTSKEVER, I.; TALWAR, K.; TUCKER, P.; VANHOUCKE, V.; VASUDEVAN, V.; VIéGAS, F.; VINYALS, O.; WARDEN, P.; WATTENBERG, M.; WICKE, M.; YU, Y.; ZHENG, X. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <https://www.tensorflow.org/>.

AGARWAL, V.; ABIDI, B. R.; KOSCHAN, A.; ABIDI, M. A. An overview of color constancy algorithms. *Journal of Pattern Recognition Research*, v. 1, n. 1, p. 42–54, 2006.

ALEXE, B.; DESELAERS, T.; FERRARI, V. What is an object? In: IEEE. *2010 IEEE computer society conference on computer vision and pattern recognition*. [S.l.], 2010. p. 73–80.

ARBELAEZ, P.; MAIRE, M.; FOWLKES, C.; MALIK, J. Contour detection and hierarchical image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 33, n. 5, p. 898–916, 2010.

BARNARD, K.; MARTIN, L.; COATH, A.; FUNT, B. A comparison of computational color constancy algorithms-part ii: Experiments with image data. *IEEE transactions on Image Processing*, v. 11, n. 9, p. 985–996, 2002.

BARRON, J. T. Convolutional color constancy. In: *Proceedings of the IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2015. p. 379–387.

BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 35, n. 8, p. 1798–1828, 2013.

BEZERRA, E. Introdução à aprendizagem profunda. *Artigo–31º Simpósio Brasileiro de Banco de Dados–SBBD2016–Salvador*, 2016.

BIANCO, S.; CUSANO, C.; SCHETTINI, R. Color constancy using cnns. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. [S.l.: s.n.], 2015. p. 81–89.

BIANCO, S.; CUSANO, C.; SCHETTINI, R. Single and multiple illuminant estimation using convolutional neural networks. *IEEE Transactions on Image Processing*, IEEE, v. 26, n. 9, p. 4347–4362, 2017.

BORJI, A.; CHENG, M.; JIANG, H.; LI, J. Salient object detection: A survey. corr. *arXiv preprint arXiv:1411.5878*, 2014.

BUCHSBAUM, G. A spatial processor model for object colour perception. *Journal of the Franklin institute*, Elsevier, v. 310, n. 1, p. 1–26, 1980.

CELIK, T.; TJAHJADI, T. Contextual and variational contrast enhancement. *IEEE Transactions on Image Processing*, IEEE, v. 20, n. 12, p. 3431–3441, 2011.

CHEN, Q.; XU, J.; KOLTUN, V. Fast image processing with fully-convolutional networks. In: *Proceedings of the IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2017. p. 2497–2506.

CHENG, D.; PRASAD, D. K.; BROWN, M. S. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, Optical Society of America, v. 31, n. 5, p. 1049–1058, 2014.

CHENG, Y. Mean shift, mode seeking, and clustering. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 17, n. 8, p. 790–799, 1995.

DABOV, K.; FOI, A.; KATKOVNIK, V.; EGIAZARIAN, K. Image restoration by sparse 3d transform-domain collaborative filtering. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Image Processing: Algorithms and Systems VI*. [S.l.], 2008. v. 6812, p. 681207.

DENG, J.; DONG, W.; SOCHER, R.; LI, L.-J.; LI, K.; FEI-FEI, L. Imagenet: A large-scale hierarchical image database. In: IEEE. *2009 IEEE conference on computer vision and pattern recognition*. [S.l.], 2009. p. 248–255.

DICARLO, J. J.; ZOCCOLAN, D.; RUST, N. C. How does the brain solve visual object recognition? *Neuron*, Elsevier, v. 73, n. 3, p. 415–434, 2012.

DONG, X.; WANG, G.; PANG, Y.; LI, W.; WEN, J.; MENG, W.; LU, Y. Fast efficient algorithm for enhancement of low lighting video. In: IEEE. *2011 IEEE International Conference on Multimedia and Expo*. [S.l.], 2011. p. 1–6.

EBNER, M. *Color constancy*. [S.l.]: John Wiley & Sons, 2007. v. 7.

FORSYTH, D. A. A novel algorithm for color constancy. *International Journal of Computer Vision*, Springer, v. 5, n. 1, p. 5–35, 1990.

FRANZEN, R. *Kodak Lossless True Color Image Suite*. 1999. Disponível em: <http://r0k.us/graphics/kodak/>. Acessado em 22 de Janeiro de 2020.

FU, X.; ZENG, D.; HUANG, Y.; ZHANG, X.-P.; DING, X. A weighted variational model for simultaneous reflectance and illumination estimation. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2016. p. 2782–2790.

GIJSENIJ, A.; GEVERS, T.; WEIJER, J. V. D. Computational color constancy: Survey and experiments. *IEEE Transactions on Image Processing*, IEEE, v. 20, n. 9, p. 2475–2489, 2011.

GOMES, A. *Porque nós enxergamos as cores sempre iguais: entendendo o fenômeno da constância*. 2017. Disponível em: <https://www.forebrain.com.br/noticias/porque-nos-enxagamos-as-cores-sempre-iguais-entendendo-o-fenomedo-de-constancia/>. Acessado em 04 de Fevereiro de 2020.

GONZALEZ, R. C.; WOODS, R. E. *Digital Image Processing (3rd Edition)*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 2006. ISBN 013168728X.

GRAND, Y. L. *Light, colour, and vision*. [S.l.]: Wiley, 1957.

GUO, X.; LI, Y.; LING, H. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing*, IEEE, v. 26, n. 2, p. 982–993, 2016.

HORDLEY, S. D.; FINLAYSON, G. D. Re-evaluating colour constancy algorithms. In: IEEE. *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004.* [S.l.], 2004. v. 1, p. 76–79.

HOU, X.; ZHANG, L. Saliency detection: A spectral residual approach. In: IEEE. *2007 IEEE Conference on computer vision and pattern recognition*. [S.l.], 2007. p. 1–8.

HU, Y.; WANG, B.; LIN, S. Fc4: Fully convolutional color constancy with confidence-weighted pooling. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. [S.l.: s.n.], 2017. p. 4085–4094.

HU, Y.; WANG, B.; LIN, S. S. *Fully convolutional color constancy with confidence weighted pooling*. [S.l.]: Google Patents, 2019. US Patent App. 15/649,950.

IANDOLA, F. N.; HAN, S.; MOSKEWICZ, M. W.; ASHRAF, K.; DALLY, W. J.; KEUTZER, K. Squeezenet: Alexnet-level accuracy with 50x fewer parameters and< 0.5 mb model size. *arXiv preprint arXiv:1602.07360*, 2016.

JIANG, B.; ZHANG, L.; LU, H.; YANG, C.; YANG, M.-H. Saliency detection via absorbing markov chain. In: *Proceedings of the IEEE international conference on computer vision*. [S.l.: s.n.], 2013. p. 1665–1672.

JOBSON, D. J.; RAHMAN, Z.-u.; WOODELL, G. A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image processing*, IEEE, v. 6, n. 7, p. 965–976, 1997.

KOSCHAN, A.; ABIDI, M. *Digital color image processing*. [S.l.]: John Wiley & Sons, 2008.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2012. p. 1097–1105.

LAND, E. H. The retinex theory of color vision. *Scientific american*, JSTOR, v. 237, n. 6, p. 108–129, 1977.

LAND, E. H.; MCCANN, J. J. Lightness and retinex theory. *Josa*, Optical Society of America, v. 61, n. 1, p. 1–11, 1971.

LEE, C.; LEE, C.; KIM, C.-S. Contrast enhancement based on layered difference representation. In: IEEE. *2012 19th IEEE International Conference on Image Processing*. [S.l.], 2012. p. 965–968.

LEE, C.; LEE, C.; KIM, C.-S. Contrast enhancement based on layered difference representation of 2d histograms. *IEEE transactions on image processing*, IEEE, v. 22, n. 12, p. 5372–5384, 2013.

LI, C.; GUO, J.; PORIKLI, F.; PANG, Y. Lightennet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognition Letters*, Elsevier, v. 104, p. 15–22, 2018.

LI, X.; LU, H.; ZHANG, L.; RUAN, X.; YANG, M.-H. Saliency detection via dense and sparse reconstruction. In: *Proceedings of the IEEE International Conference on Computer Vision*. [S.l.: s.n.], 2013. p. 2976–2983.

LORE, K. G.; AKINTAYO, A.; SARKAR, S. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, Elsevier, v. 61, p. 650–662, 2017.

LOU, Z.; GEVERS, T.; HU, N.; LUCASSEN, M. P. et al. Color constancy by deep learning. In: *BMVC*. [S.l.: s.n.], 2015. p. 76–1.

LV, F.; LU, F.; WU, J.; LIM, C. Mbllen: Low-light image/video enhancement using cnns. In: *BMVC*. [S.l.: s.n.], 2018. p. 220.

MA, K.; ZENG, K.; WANG, Z. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing*, IEEE, v. 24, n. 11, p. 3345–3356, 2015.

MARTIN, D.; FOWLKES, C.; TAL, D.; MALIK, J. et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV VANCOUVER:. [S.l.], 2001.

MELLO, C. A.; SARAIVA, M. M.; MENOR, D. P.; NISHIHARA, R. A comparative study of objective video quality assessment metrics. *J. UCS*, v. 23, n. 5, p. 505–527, 2017.

MURRAY, N.; VANRELL, M.; OTAZU, X.; PARRAGA, C. A. Saliency estimation using a non-parametric low-level vision model. In: IEEE. *CVPR 2011*. [S.l.], 2011. p. 433–440.

NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: *Proceedings of the 27th international conference on machine learning (ICML-10)*. [S.l.: s.n.], 2010. p. 807–814.

NETTO, A. V. *Processamento e análise de imagens para medição de vícios de refração ocular*. Tese (Doutorado) — Universidade de São Paulo, 2003.

OH, S. W.; KIM, S. J. Approaching the computational color constancy as a classification problem through deep learning. *Pattern Recognition*, Elsevier, v. 61, p. 405–416, 2017.

PAULA, N. G. d. et al. *Detecção de região saliente em imagens usando dissimilaridade de cor e amostragem por pixeis aleatórios*. Dissertação (Mestrado) — Universidade Tecnológica Federal do Paraná, 2015.

PITKÄNEN, P. Automatic image quality enhancement using deep neural networks. 2019.

PIZER, S. M. Contrast-limited adaptive histogram equalization: Speed and effectiveness stephen m. pizer, r. eugene johnston, james p. ericksen, bonnie c. yankaskas, keith e. muller medical image display research group. In: IEEE COMPUTER SOCIETY PRESS. *Proceedings of the First Conference on Visualization in Biomedical Computing, Atlanta, Georgia, May 22-25, 1990*. [S.l.], 1990. p. 337.

PONTI, M. A.; COSTA, G. B. P. da. Como funciona o deep learning. *arXiv preprint arXiv:1806.07908*, 2018.

RAHMAN, Z.; AAMIR, M.; PU, Y.-F.; ULLAH, F.; DAI, Q. A smart system for low-light image enhancement with color constancy and detail manipulation in complex light environments. *Symmetry*, Multidisciplinary Digital Publishing Institute, v. 10, n. 12, p. 718, 2018.

RAHTU, E.; KANNALA, J.; SALO, M.; HEIKKILÄ, J. Segmenting salient objects from images and videos. In: SPRINGER. *European conference on computer vision*. [S.l.], 2010. p. 366–379.

REZA, A. M. Realization of the contrast limited adaptive histogram equalization (clahe) for real-time image enhancement. *Journal of VLSI signal processing systems for signal, image and video technology*, Springer, v. 38, n. 1, p. 35–44, 2004.

SCHMIDHUBER, J. Deep learning in neural networks: An overview. *Neural networks*, Elsevier, v. 61, p. 85–117, 2015.

SEO, H. J.; MILANFAR, P. Static and space-time visual saliency detection by self-resemblance. *Journal of vision*, The Association for Research in Vision and Ophthalmology, v. 9, n. 12, p. 15–15, 2009.

SHI, L. Re-processed version of the gehler color constancy dataset of 568 images. *http://www. cs. sfu. ca/˜ color/data/*, 2000.

SHI, W.; LOY, C. C.; TANG, X. Deep specialized network for illuminant estimation. In: SPRINGER. *European Conference on Computer Vision*. [S.l.], 2016. p. 371–387.

TAO, L.; ZHU, C.; XIANG, G.; LI, Y.; JIA, H.; XIE, X. Llcnn: A convolutional neural network for low-light image enhancement. In: IEEE. *2017 IEEE Visual Communications and Image Processing (VCIP)*. [S.l.], 2017. p. 1–4.

TIAN, Q.-C. *Color Correction and Contrast Enhancement for Natural Images and Videos*. Tese (Doutorado), 2018.

VONIKAKIS, V.; KOUSKOURIDAS, R.; GASTERATOS, A. On the evaluation of illumination compensation algorithms. *Multimedia Tools and Applications*, Springer, v. 77, n. 8, p. 9211–9231, 2018.

WANG, S.; ZHENG, J.; HU, H.-M.; LI, B. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Transactions on Image Processing*, IEEE, v. 22, n. 9, p. 3538–3548, 2013.

WANG, Z.; BOVIK, A. C.; SHEIKH, H. R.; SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, IEEE, v. 13, n. 4, p. 600–612, 2004.

WEIJER, J. V. D.; GEVERS, T.; GIJSENIJ, A. Edge-based color constancy. *IEEE Transactions on image processing*, IEEE, v. 16, n. 9, p. 2207–2214, 2007.

WEIJER, J. Van de; GEVERS, T.; BAGDANOV, A. D. Boosting color saliency in image feature detection. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 28, n. 1, p. 150–156, 2005.

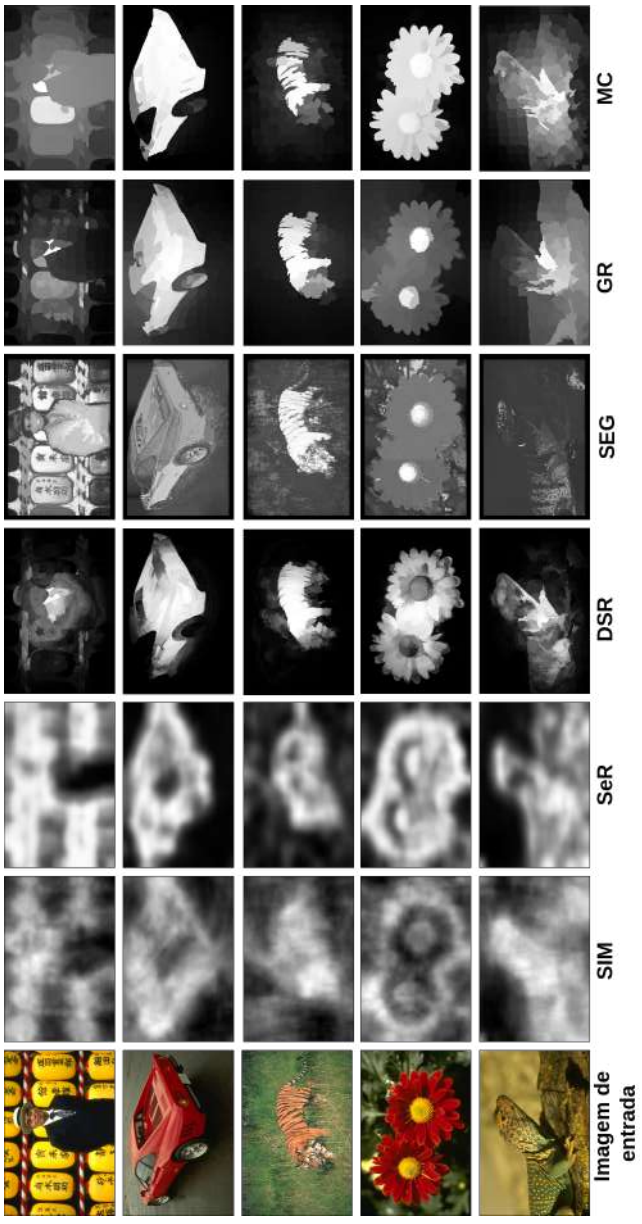
YANG, C.; ZHANG, L.; LU, H. Graph-regularized saliency detection with convex-hull-based center prior. *IEEE Signal Processing Letters*, IEEE, v. 20, n. 7, p. 637–640, 2013.

YING, Z.; LI, G.; GAO, W. A bio-inspired multi-exposure fusion framework for low-light image enhancement. *arXiv preprint arXiv:1711.00591*, 2017.

ZHANG, K.; ZUO, W.; CHEN, Y.; MENG, D.; ZHANG, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, IEEE, v. 26, n. 7, p. 3142–3155, 2017.

APÊNDICE A – EXEMPLOS DE IMAGENS UTILIZADAS PARA ESCOLHA DE MÉTODO DETECTOR DE SALIÊNCIA

Figura 52 – Comparação entre mapas de saliência gerados pelos métodos SIM, SeR, DSR, SEG, GR e MC.



Fonte: O autor.