



Joel Felipe de Oliveira Gaya

Processamento de Imagens Adquiridas em Meios Participativos Usando Deep Learning

Rio Grande, 2018

Joel Felipe de Oliveira Gaya

Processamento de Imagens Adquiridas em Meios Participativos Usando Deep Learning

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande - FURG

Universidade Federal do Rio Grande - FURG

Centro de Ciências Computacionais - C3

Programa de Pós-Graduação em Computação

Orientadora: Prof^a. Dr^a. Silvia Silva da Costa Botelho

Coorientador: Prof. Dr. Paulo Lilles Jorge Drews Junior

Rio Grande,

2018

Ficha catalográfica

G285p Gaya, Joel Felipe de Oliveira.
Processamento de imagens adquiridas em meios participativos usando
Deep Learning /Joel Felipe de Oliveira Gaya. – 2018.
119 p.

Dissertação (mestrado) – Universidade Federal do Rio Grande – FURG,
Programa de Pós-graduação em Computação, Rio Grande/RS, 2018.
Orientadora: Dr^a. Sílvia Silva da Costa Botelho.
Coorientador: Dr. Paulo Lilles Jorge Drews Junior.

1. Processamento de imagens 2. *Deep Learning* 3. Aprendizado de
máquina 4. Dehazing I. Botelho, Sílvia Silva da Costa II. Drews Junior,
Paulo Lilles Jorge III. Título.

CDU 004.92



MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DO RIO GRANDE
CENTRO DE CIÊNCIAS COMPUTACIONAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO
CURSO DE MESTRADO EM ENGENHARIA DE COMPUTAÇÃO

DISSERTAÇÃO DE MESTRADO

Processamento de Imagens Adquiridas em Meios Participativos Usando
Deep Learning

Joel Felipe de Oliveira Gaya

Banca examinadora:

Prof. Dr. Carlos Rodrigues Rocha

Prof. Dr. Ricardo Nagel Rodrigues

Profª. Drª. Graçaliz Pereira Dimuro

Profª. Drª. Silvia Silva da Costa Botelho
Orientadora

Prof. Dr. Paulo Lilles Jorge Drews Junior
Coorientador

Resumo

Este trabalho aborda o problema dos efeitos causados pela interação dos raios de luz com as partículas de um meio participativo em imagens ópticas digitais. A abordagem adotada utiliza uma rede neural que recebe uma imagem com influência do meio e é treinada para produzir uma imagem correspondente sem esta influência. Para que o modelo resultante seja capaz de realizar esta tarefa preservando as estruturas da imagem de entrada, é proposta a utilização do *guided filter* como parte da rede.

O conjunto de dados utilizado no treinamento é composto por imagens internas com degradação simulada através de um modelo simplificado de formação da imagem. Entretanto, a rede adotada não utiliza explicitamente este modelo, permitindo que simulações menos simplificadas sejam utilizadas futuramente sem que a rede precise ser alterada.

Os resultados obtidos através desta abordagem são avaliados em relação ao estado-da-arte de forma qualitativa, utilizando imagens externas reais, e de forma quantitativa, utilizando o *dataset* D-Hazy. De acordo com a análise quantitativa, os resultados obtidos alcançaram o segundo lugar neste *dataset* para a porção de imagens derivadas do *Middlebury* dataset.

Palavras-chaves: Processamento de Imagens; *Deep Learning*; Aprendizado de Máquina; *Dehazing*

Abstract

This work addresses the problem of the effects caused by the interaction between particles and light rays in a participatory medium in digital optical images. Our approach uses a neural network that receives an image with midium influence and is trained to produce a corresponding image without this effects. In order to allow the resulting model to perform this task while preserving the input image structures, we propose the guided filter use as part of the network.

Our dataset is composed by indoor images with simulated degradation through a simplified model of image formation. However, the network does not explicitly use this model, allowing more complex simulations to be used in the future without requiring any network change.

The results obtained through this approach are qualitatively compared to state-of-the-art, using real outdoor images, and quantitatively using the D-Hazy dataset. According to the quantitative analysis, the obtained results reached the second place in this dataset for the images derived from the *Middlebury* dataset.

Key-words: Image Processing; Deep Learning; Machine Learning; Dehazing

Lista de ilustrações

Figura 1 – Representação da metodologia utilizada para o treinamento da rede. . .	11
Figura 2 – Representação das principais interações dos raios luminosos com as partículas em suspensão.	15
Figura 3 – Componentes do processo de formação da imagem em meios participativos.	16
Figura 4 – Trajetória dos raios que compõem a componente direta, partindo da fonte de luz (representada pelo sol) até o observador (olho).	17
Figura 5 – Trajetória dos raios de luz que formam o <i>backscattering</i>	18
Figura 6 – Modelo simplificado de propagação da luz em meios participativos. . .	20
Figura 7 – Representação da metodologia utilizada pelo método proposto.	25
Figura 8 – Estrutura de rede utilizada.	29
Figura 9 – Estrutura utilizada para rede discriminadora.	31
Figura 10 – Exemplos de imagens utilizadas no simulador.	33
Figura 11 – Exemplos de simulação utilizando parâmetros para névoa e para água. . .	36
Figura 12 – Comparativo entre a Dehazenet e sua versão que utiliza o <i>guided filter</i> durante o treinamento.	37
Figura 13 – Comparativo entre a MSCNN e sua versão que utiliza o <i>guided filter</i> durante o treinamento.	38
Figura 14 – Comparativo entre a rede utilizada neste trabalho (identificada como Baseada na Unet + Guided Filter), sua versão sem <i>guided filter</i> (identificada como Baseada na Unet) , e sua versão sem <i>guided filter</i> e sem <i>skip connections</i> (Encoder-Decoder).	39
Figura 15 – Exemplos de imagens presentes no <i>dataset</i> D-Hazy provenientes do <i>Middlebury</i>	40
Figura 16 – Comparativo utilizando a imagem Logos.	44
Figura 17 – Comparativo utilizando a imagem Cityscape.	45
Figura 18 – Comparativo utilizando a imagem Girls.	46
Figura 19 – Comparativo utilizando a imagem Wheat Field	46
Figura 20 – Comparativo utilizando a imagem Tiananmen.	47
Figura 21 – Comparativo utilizando a imagem Peixes.	47
Figura 22 – Comparativo utilizando a imagem Mergulhadores.	48

Lista de tabelas

Tabela 1 – Parâmetros utilizados no simulador para névoa.	35
Tabela 2 – Parâmetros utilizados no simulador para água.	36
Tabela 3 – Comparativo entre as funções de erro testadas usando imagens da porção do <i>dataset</i> D-Hazy derivada do <i>dataset</i> Middlebury.	41
Tabela 4 – Avaliação quantitativa do método proposto usando imagens da porção do <i>dataset</i> D-Hazy derivada do <i>dataset</i> Middlebury, usando o SSIM. . .	42
Tabela 5 – Avaliação quantitativa do método proposto usando imagens da porção do <i>dataset</i> D-Hazy derivada do <i>dataset</i> Middlebury, usando o CI-EDE2000.	43

Lista de Abreviaturas e Siglas

CNN *Convolutional Neural Network* - Rede Neural Convolucional

EAM Erro Absoluto Médio

FR-IQA *Full-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Completa

GPU *Graphic Processing Unit* - Unidade de Processamento Gráfico

GAN *Generative Adversarial Nets* - Redes Adversárias Generativas

IQA *Image Quality Assessment* - Avaliação da Qualidade de Imagem

LOS *Line of Sight* - Linha de Visada

NR-IQA *No-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem sem Imagem de Referência

ROV *Remotely Operated Vehicle* - Veículo Operado Remotamente

RR-IQA *Reduced-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Reduzida

Sumário

1	Introdução	10
1.1	Motivação	11
1.2	Objetivos	13
1.3	Organização do texto	13
2	Referencial Teórico	15
2.1	Modelo de formação de imagens ópticas em Meios Participativos	15
2.2	Métodos para Redução da Turbidez em Imagens	21
3	Uma Metodologia para Redução da Turbidez em Imagens Ópticas	25
3.1	Guided Filter	26
3.2	Rede Neural para Redução da Turbidez em Imagens	28
3.3	Funções de Erro	29
3.4	Dados usados no treinamento	31
4	Resultados Experimentais	34
4.1	Avaliação das Decisões Tomadas em Relação à Estrutura da Rede	35
4.2	Análise Quantitativa Utilizando Imagens Internas Sintéticas	38
4.2.1	Avaliação do Uso de Diferentes Funções de Erro	39
4.2.2	Avaliação Quantitativa em Relação ao Estado-da-arte	40
4.3	Análise Qualitativa Utilizando Imagens Reais Externas	44
5	Considerações Finais e Trabalhos Futuros	49
	Referências	51

1 Introdução

Imagens capturadas em um ambiente externo frequentemente sofrem a influência de partículas suspensas no ar. Estas partículas interagem com os raios de luz e causam *espalhamento* e *absorção*. Mesmo em dias ensolarados e com pouca influência de partículas em suspensão, pode-se observar que objetos mais distantes tendem a ser percebidos como mais esbranquiçados e com pouco contraste. Em dias de névoa, estes efeitos podem reduzir a visibilidade de forma drástica.

Ao adquirir imagens em ambientes subaquáticos, obtém-se efeitos bastante semelhantes. Porém, diferente da névoa, a água absorve determinados comprimentos de onda mais do que outros. Mais especificamente, em regiões com pouca matéria orgânica e sedimentos dissolvidos na água, o vermelho é mais absorvido, seguido pelo verde, fazendo então com que a cor azul predomine. Em regiões com muita concentração de plâncton, o azul pode ser bastante absorvido pela matéria orgânica, e com isso os raios de luz com comprimento de onda próximos ao verde acabam sendo os menos absorvidos (ANCUTI et al., 2017).

Neste trabalho, os efeitos causados por essas interações da luz com o meio no qual ela se propaga serão denominados *turbidez*, visto que este é o nome de uma propriedade física dos fluidos, que se traduz na dificuldade da passagem da luz devido à presença de materiais em suspensão. Portanto, o termo da língua portuguesa que mais se aproxima dos efeitos aqui tratados.

O problema abordado neste trabalho é o da remoção da turbidez em imagens adquiridas em meios participativos utilizando apenas as informações presentes na imagem. Um meio é participativo se as partículas presentes nele interagem com a luz, este trabalho tem como foco a água e a névoa. Para tratar esse problema optou-se pelo uso de técnicas de *deep learning*. Esta escolha foi baseada no bom desempenho destas técnicas em diversas tarefas de processamento de imagens (LEDIG et al., 2016; CHEN et al., 2016). *Deep learning* é um ramo de aprendizado de máquina baseado em modelos compostos por muitas camadas de processamento, que realizam transformações não lineares sobre os dados. Neste contexto, uma rede neural é uma composição destas camadas.

Devido à natureza dos efeitos causados por meios participativos, adquirir um conjunto de dados com pares de imagens da mesma cena, sob as mesmas condições de iluminação, ângulo de câmera e onde uma das imagens tenha sido capturada com um meio participativo e outra sem ele é extremamente difícil. Por esta razão decidiu-se treinar a rede neural utilizando um conjunto de dados simulados. A Figura 1 apresenta uma ilustração da metodologia adotada. Nesta metodologia, são utilizadas imagens coloridas

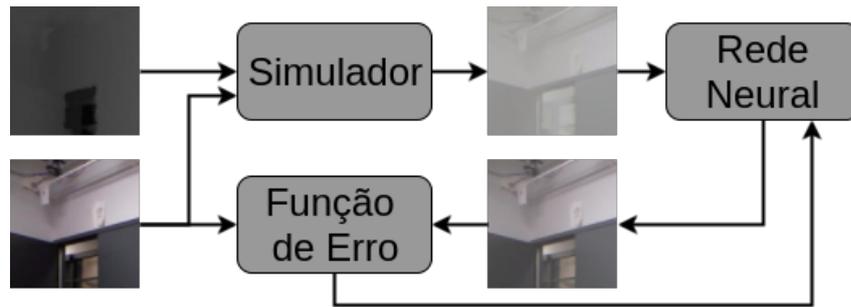


Figura 1: Representação da metodologia utilizada para o treinamento da rede. O simulador recebe como entrada uma imagem colorida e seu respectivo mapa de profundidade (distância de cada pixel à cena). Como saída o simulador gera uma imagem com turbidez. A imagem produzida pelo simulador é a entrada da rede neural, que tem o objetivo de gerar uma imagem próxima à imagem de referência (entrada do simulador). A função de erro tem o papel de receber a imagem de referência e a saída da rede e por meio de algum critério definir o quão distintas elas são.

capturadas em ambientes internos e a distância de cada pixel à cena para produzir uma imagem com turbidez simulada. A imagem turva é então utilizada como entrada de uma rede neural treinada para produzir uma captura desta mesma cena sem turbidez.

1.1 Motivação

Os efeitos causados por meios participativos no processo de formação de imagens dificultam uma série de tarefas dependentes de visão. Sejam tarefas que envolvam visão computacional, como segmentação (CHEN et al., 2016) e classificação (SZEGEDY et al., 2016), ou desempenhadas por um operador humano, como dirigir em condições de nevoeiro ou em operações envolvendo *Remotely Operated Vehicle* - Veículo Operado Remotamente (ROV).

A presença de névoa pode ocasionar transtornos referentes ao trânsito de veículos terrestres, aeronaves e embarcações. Sua presença pode causar drástica diminuição na segurança e por vezes impossibilita a operação em portos e aeroportos, o que representa perdas econômicas.

Ambientes como oceanos e mares são interessantes para exploração, pois possuem grande extensão, biodiversidade e variedade de minerais. Entretanto, sua exploração ainda envolve uma série de desafios, dentre eles o impacto decorrente da dificuldade de percepção utilizando sensores ópticos em uma série de tarefas (CODEVILLA et al., 2015).

Neste contexto, a obtenção de imagens com menor influência do meio através de um método de restauração assume relevância e valor econômico. Seja como uma etapa de pré-processamento para algoritmos de visão computacional que desprezam a influência do meio ou até mesmo para que um especialista humano possa examinar essas imagens de uma forma mais eficiente.

Explorando o aumento de visibilidade em imagens degradadas por névoa, tempestade de areia e *smog* industrial, veículos terrestres podem passar a trafegar com mais segurança nas estradas mesmo sob estas condições desfavoráveis (ROSENBLUM; GOTHARD, 2000; TAREL; HAUTIERE, 2009). Trens e portos podem também ser beneficiados, já que a visibilidade é importante para garantir a segurança de trabalhadores e passageiros. Além disso, a obtenção de imagens com menos turbidez em ambientes subaquáticos aumenta a segurança em operações com ROVs, possibilitando que o robô mantenha uma distância maior do objeto observado, reduzindo o risco de uma colisão (JASIOBEDZKI et al., 2008).

Uma série de trabalhos vem sendo realizados a fim de tratar imagens degradadas por meios participativos. Eles normalmente são classificados de acordo com sua abordagem como métodos de *realce* e *restauração*.

Métodos de realce tem como objetivo melhorar a qualidade da imagem. Seja a melhora da qualidade subjetiva (visual) ou através de alguma métrica, como contraste, por exemplo. Nestes métodos, o modelo físico do fenômeno de propagação da luz não é necessário. Exemplos de métodos de realce são (BAZEILLE et al., 2006; ANCUTI et al., 2012; HUMMEL, 1977).

Métodos de restauração se concentram em resolver a equação do fenômeno de propagação da luz, a fim de recuperar a imagem sem a influência do meio. Dada a dificuldade em estimar os parâmetros da equação do meio, diversos modelos mais simplificados são utilizados. São considerados métodos de restauração (SCHECHNER; KARPEL, 2005; FATTAL, 2008; GALDRAN et al., 2015; HE; SUN; TANG, 2011; CAI et al., 2016a; CAI et al., 2016b; BERMAN; TREIBITZ; AVIDAN, 2016), entre outros.

Abordagens recentes que utilizam aprendizado de máquina podem também ser classificadas como métodos de restauração. Estes utilizam as técnicas de aprendizado de máquina para extrair informações que são posteriormente utilizadas no modelo físico do fenômeno, como por exemplo em Cai et al. (2016a), Cai et al. (2016b) e Li et al. (2017).

A solução aqui proposta não procura explicitamente resolver a equação do fenômeno. Ao invés disso, uma rede neural é treinada para a partir de uma imagem com turbidez produzir uma saída correspondente sem a degradação. No entanto, esta abordagem se baseia no modelo físico para gerar as imagens usadas durante a etapa de treinamento, apresentando-se como uma abordagem híbrida entre realce e restauração. Este método, além de permitir simulações mais complexas e realistas que as utilizadas pelos métodos restauração, se beneficia de melhorias na simulação sem que a estrutura da rede precise ser modificada.

Vale ressaltar que um desafio relacionado ao problema de remoção da turbidez está ligado à variação espacial na intensidade do fenômeno. A utilização de redes neurais, sem considerar esta variação espacial, pode causar a perda de estruturas da imagem. Com isso, é interessante a utilização de mecanismos capazes de manter ou recuperar estruturas perdidas ao longo da rede.

1.2 Objetivos

O objetivo principal deste trabalho é desenvolver uma metodologia que utilize redes neurais profundas para reconstrução de imagens degradadas por meios participativos, que leve em consideração a dificuldade de manter as estruturas presentes na imagem após o processamento. Para tal o trabalho apresenta os seguintes objetivos específicos:

- Estudar o fenômeno de propagação da luz em meios participativos;
- Fazer uma revisão dos métodos presentes no estado-da-arte;
- Criar um conjunto de dados simulados para treinar a rede neural;
- Definir uma estrutura de rede capaz de realizar as operações necessárias para transformar a imagem degradada em sua versão sem influência do meio;
- Avaliar o uso de diferentes funções de erro durante o treinamento;
- Fazer a análise quantitativa do método em relação ao estado-da-arte usando imagens sintéticas.
- Fazer a análise comparativa do método em relação ao estado-da-arte de forma qualitativa usando imagens reais;

1.3 Organização do texto

Além do presente capítulo, este trabalho está organizado da seguinte maneira:

- O [Capítulo 2](#) introduz a fundamentação teórica acerca do problema abordado nesta dissertação e uma breve apresentação dos trabalhos relacionados;
- O [Capítulo 3](#) apresenta a descrição da abordagem utilizada no desenvolvimento deste trabalho;

- O [Capítulo 4](#) apresenta o impacto de algumas decisões tomadas em relação a estrutura da rede, um comparativo entre diferentes funções de erro na metodologia proposta e os resultados obtidos pelo método proposto em comparação ao estado-da-arte;
- O [Capítulo 5](#) traz as considerações finais e sugestões de trabalhos futuros.

2 Referencial Teórico

Neste capítulo serão apresentados conceitos importantes para o entendimento do problema abordado e das decisões tomadas durante o desenvolvimento do presente trabalho. A [Seção 2.1](#) traz a modelagem dos efeitos observados nas imagens capturadas em meios participativos. Esta modelagem além de ser utilizada na simulação dos dados para o treinamento da rede proposta neste trabalho, é fundamental para uma série de métodos que procuram reduzir a turbidez em imagens. A [Seção 2.2](#) apresenta as abordagens adotadas por métodos que tem o objetivo de reduzir a turbidez em imagens ópticas.

2.1 Modelo de formação de imagens ópticas em Meios Participativos

Quando um raio de luz se propaga através de um meio participativo, ele pode interagir com as partículas em suspensão. As formas de interação mais significativas são o *espalhamento* e a *absorção*, ilustrados na [Figura 2](#). A absorção e o espalhamento causam uma perda de informação visual. A parcela perdida representa uma limitação física para a recuperação do sinal, pois este não alcança seu destino.

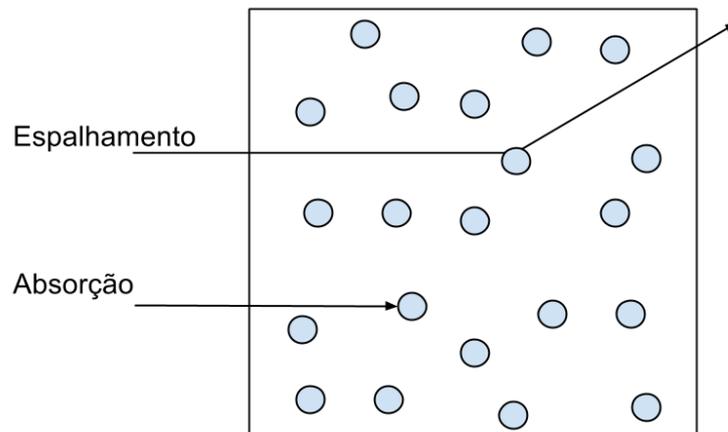


Figura 2: Representação das principais interações dos raios luminosos com as partículas em suspensão. Na *absorção* o raio de luz é absorvido por uma das partículas do meio, enquanto no *espalhamento* o raio de luz é desviado de sua trajetória retilínea.

Considerando um feixe de inicialmente n_0 fótons, que se propaga em um meio homogêneo com coeficientes de espalhamento e absorção conhecidos, a perda de informação pode ser descrita com base na *lei da Atenuação de Lambert*, [Equação 2.1](#) ([JAFPE, 1990](#)).

$$n = n_0 e^{-cd}, \quad (2.1)$$

sendo n o número de fótons que chegam ao seu destino, d a distância percorrida pelo feixe e c a soma dos coeficientes de atenuação e de espalhamento.

O espalhamento, além de reduzir a intensidade de um feixe, também é responsável por injetar energia em pontos que não deveriam recebê-la. A informação luminosa adicionada pelo espalhamento produz dois fenômenos na imagem capturada. São eles, *Forward Scattering* e *Backscattering*.

No primeiro, os raios de luz vindos da cena observada são espalhados pelo meio em ângulos pequenos fazendo com que alguns fótons atinjam pontos vizinhos ao que atingiriam originalmente. Ele é responsável por criar um borramento na imagem formada.

No segundo, raios de luz que não são provenientes da cena observada são desviados por partículas do meio, para o plano da imagem criando uma espécie de véu que cobre a imagem resultante.

O modelo de formação da imagens em meios participativos pode ser descrito por McGlamery (1980) e Jaffe (1990). Neste, a irradiância em cada pixel da imagem capturada pode ser descrita pela soma de três componentes (Equação 2.2), conforme ilustrado na Figura 3.

$$E_T = E_d + E_f + E_{bs}, \quad (2.2)$$

onde E_T é a irradiação total em um pixel, E_d é a componente direta, E_f é a componente que representa o *forward scattering* e E_{bs} é a componente que compreende o *backscattering*.

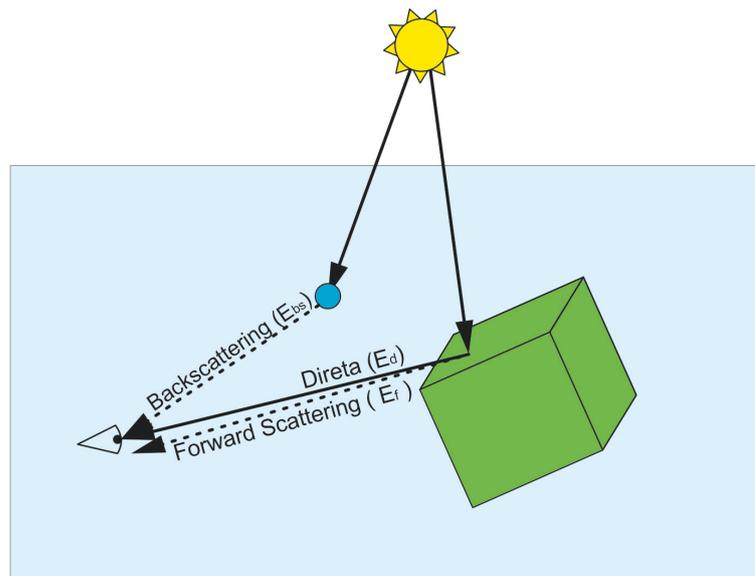


Figura 3: Representação das três componentes do processo de formação de imagem em meios participativos: Componente Direta, Forward Scattering e Backscattering. Na cena representada, o observador (olho) e o objeto (cubo verde) estão imersos em um meio participativo (retângulo azul).

Schechner e Karpel (2005) mostram que nos meios participativos considerados neste trabalho, o principal fator que causa perda de contraste e visibilidade é o *backscattering*, e que a componente *forward scattering* pode ser desconsiderada, pois seu efeito é pouco expressivo na imagem resultante.

Na componente direta, ilustrada na Figura 4, a luz produzida pela fonte de iluminação segue em todas as direções e portanto a luz se propaga no meio reduzindo sua irradiância de acordo com a área da esfera centrada no ponto onde se encontra a fonte de iluminação. Assim, a diminuição se dá com o quadrado da distância até a fonte luminosa.

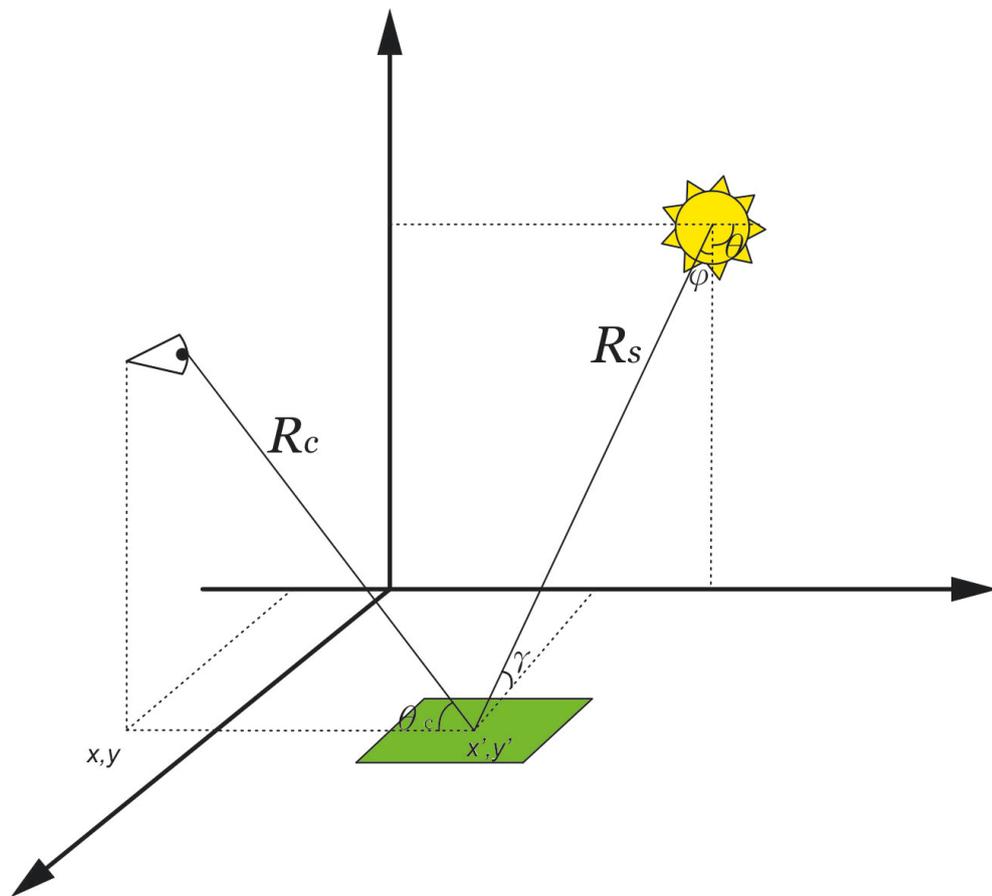


Figura 4: Trajetória dos raios que compõem a componente direta, partindo da fonte de luz até o observador. Nesta ilustração θ e φ são os ângulos que indicam a direção que o raio de luz foi emitido pela fonte de iluminação, γ é o ângulo entre o raio de luz e a superfície da cena, R_c é a distância entre um ponto da cena e o observador (olho) e R_s a distância entre este mesmo ponto e a fonte de iluminação.

Como citado anteriormente a luz sofre a atenuação do meio descrita pela lei da Atenuação de Lambert, Equação 2.1, e com isso ocorre a perda de informação luminosa antes mesmo de os raios incidirem na cena. O processo da fonte até a cena é modelado pela Equação 2.3:

$$E_I(x', y') = BP(\theta, \varphi) \cos \gamma \frac{e^{-cR_s}}{R_s^2}, \quad (2.3)$$

onde $E_I(x', y')$ é a irradiância que chega à cena no ponto descrito pelas coordenadas x' e y' , cuja distância até a fonte de luz é R_s e onde a superfície forma um ângulo γ com os raios luminosos. $BP(\theta, \varphi)$ é a radiância emitida pela fonte de luz nos ângulos θ e φ .

A Equação 2.4, descreve a irradiância que chega ao observador sem ser desviada, a partir da que chega à cena.

$$E_d(x, y) = E_I(x', y', \theta, \varphi)e^{-cR_c} J(x', y'), \quad (2.4)$$

onde $J(x', y')$ é a refletância da superfície fotografada no ponto (x', y') e R_c é a distância ao observador.

Para calcular a contribuição do *backscattering*, E_{bs} , (MCGLAMERY, 1980) divide o volume entre a cena e o observador em um grande número de volumes diferenciais. A contribuição de cada volume é baseada em β , que é dependente do ângulo entre a câmera e a fonte de luz em relação ao volume, como ilustrado na Figura 5.

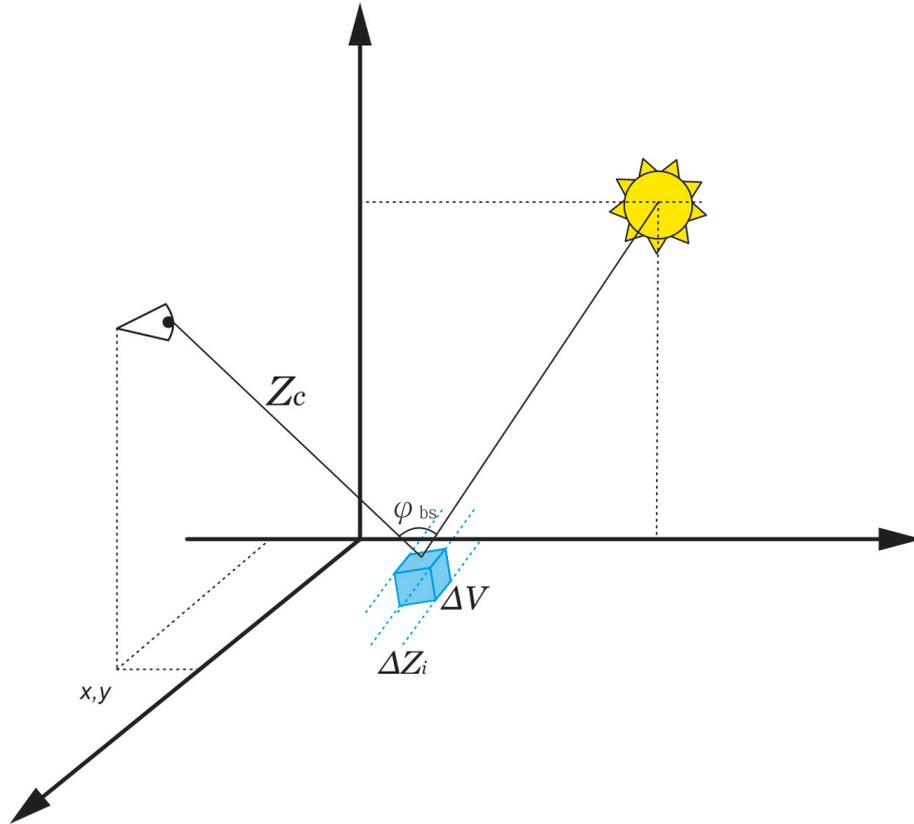


Figura 5: Representação da trajetória dos raios de luz que formam o *backscattering*. Onde um raio de luz incide em um volume infinitesimal ΔV , de comprimento ΔZ_i e é refletido em direção ao observador. Nesta ilustração Z_c é a distância entre o volume infinitesimal e a cena e φ_{bs} é a soma do ângulo de incidência com o ângulo de reflexão.

E então, é realizado o somatório da contribuição dos N volumes considerados, como mostrado na equação Equação 2.5, considerando E_s como a irradiância recebida

pelo volume que é calculada de forma semelhante à [Equação 2.3](#).

$$E_{bs}(x, y) = \sum_{i=1}^N e^{-cZc} \beta(\varphi_{bs}) E_s(x', y', z'). \quad (2.5)$$

A função $\beta(\varphi_{bs})$ é dependente das partículas do meio e modela a contribuição luminosa de um volume à partir de um ângulo φ_{bs} .

Com o modelo descrito a cima é possível simular os efeitos dos meios participativos na formação das imagens em ambientes virtuais tridimensionais. No entanto, como o simulador usado neste trabalho utiliza imagens RGB-D, onde a fonte de iluminação e parte dos objetos que compõem a cena não são conhecidos, este modelo precisa ser simplificado. Estas simplificações também são usadas no desenvolvimento de uma série de métodos de restauração de imagem ([HE; SUN; TANG, 2011](#); [FATTAL, 2014](#); [MENG et al., 2013](#)).

Considerando que a iluminação é constante para todo o volume onde a luz se propaga, o que é uma aproximação válida para grande parte dos casos de iluminação natural, é possível escrever a [Equação 2.6](#) como:

$$E_d(x, y) = L(x, y) J(x, y) e^{-cRc}, \quad (2.6)$$

onde $L(x, y)$ é a informação luminosa que chega à cena. Neste modelo apenas a atenuação do meio é considerada significativa para redução da informação luminosa proveniente da componente direta.

Faz-se necessária também algumas simplificações na componente *backscattering*. Essas simplificações foram feitas de acordo com [Schechner e Karpel \(2005\)](#). Nelas a distância é considerada até um limite z do observador e a fonte de luz ilumina os volumes da *Line of Sight* - Linha de Visada (**LOS**), com uma direção relativa $r = (\theta, \varphi)$ e intensidade I^{source} . Com isso, a contribuição desta fonte de luz pode ser calculada como:

$$B(\theta, \varphi) \approx \int_0^z k \beta(\varphi_{bs}) I^{source}(\theta, \varphi) e^{-cl} dl, \quad (2.7)$$

onde k é uma constante que depende da distância focal. Resolvendo a integral obtém-se:

$$B(\theta, \varphi) \approx A(\theta, \varphi) (1 - e^{-cz}). \quad (2.8)$$

Sendo que $A(\theta, \varphi)$ representa o *backscattering* em uma **LOS** que se estende até o infinito. A [Equação 2.9](#) representa este cenário:

$$A(\theta, \varphi) = k \beta(\varphi_{bs}) I^{source}(\theta, \varphi), \quad (2.9)$$

Neste trabalho considera-se que o *backscattering* é proveniente de uma única fonte situada no infinito. Esta consideração só pode ser feita pois assume-se que a iluminação do meio é

uniforme. Neste contexto, a integração da [Equação 2.9](#) para todos os ângulos resulta em uma constante, como mostrado na [Equação 2.10](#):

$$A = \int_{\theta, \varphi} k\beta(\varphi_{bs}) I^{source}(\theta, \varphi). \quad (2.10)$$

Por fim, o *backscattering* na cena é dado por:

$$E_{bs}(x, y) = A(1 - e^{-cz}). \quad (2.11)$$

Para essa equação, o A pode ser interpretado como a informação luminosa que vem do infinito. Ou seja, informação luminosa característica do *backscattering*.

O resultado das simplificações resultantes, é dado por:

$$I(x, y) = L(x, y)J(x, y)e^{-cz(x, y)} + A(1 - e^{-cz(x, y)}), \quad (2.12)$$

onde $J(x, y)$ representa a refletância da cena, $L(x, y)$ representa a informação luminosa que chega à cena e $z(x, y)$ a distância entre cada ponto da imagem e a cena.

As simplificações feitas sobre o modelo inicial são ilustradas na [Figura 6](#). Onde a Componente Direta é atenuada por um fator $e^{-cz(x, y)}$, denominado transmissão, e $1 - e^{-cz(x, y)}$ dá a contribuição do *backscattering* para cada feixe que parte da cena e chega ao observador.

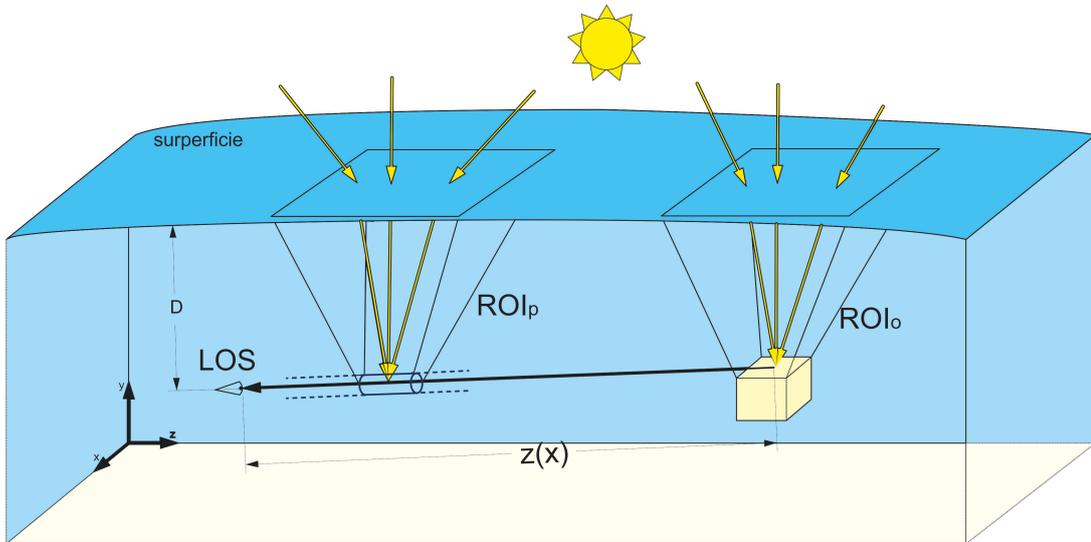


Figura 6: Modelo simplificado de propagação da luz em meios participativos. Esta é a representação de uma cena fotografada em um ambiente subaquático, onde o objeto capturado é iluminado pela luz que incide em um cone acima dele (ROI_o) e os volumes presentes na LOS são iluminados por um cone acima deles (ROI_p). Nesta ilustração, tanto o observador quanto o objeto observado estão à mesma profundidade (D) e a distância entre eles é $Z(x)$.

A [Figura 6](#) ilustra uma cena sendo fotografada em baixo d'água, com iluminação natural para uma profundidade aproximadamente constante entre o observador e a cena observada. Sendo assim, todos objetos e partículas são atingidos pela mesma informação luminosa, já que ela depende apenas da fonte de iluminação e do cone acima do objeto (ROI_o), ou acima de uma partícula (ROI_p). Onde se justifica considerar a iluminação constante tanto para a cena quanto para as partículas do meio.

Para imagens com névoa, a diferença de profundidade em que os objetos e partículas estão possui uma influência menor que na água. Isto acontece pois a turbidez costuma ser mais branda na névoa que na água. Assim, pode-se considerar que a mesma luz que ilumina a cena, também é a cor característica do *backscattering*, ou seja, $L(x, y) = A$. Logo, pode-se reescrever a [Equação 2.12](#) como:

$$I(x, y) = A J(x, y) e^{-cz(x,y)} + A(1 - e^{-cz(x,y)}). \quad (2.13)$$

É importante ressaltar que a [Equação 2.13](#) descreve o processo de formação da imagem para cada um dos canais de cor. Tendo cada um deles valores independentes em suas variáveis com exceção de $z(x, y)$, que é comum à todos eles. A [Equação 2.13](#) é posteriormente utilizada no simulador descrito na [Seção 3.4](#).

2.2 Métodos para Redução da Turbidez em Imagens

Uma série de trabalhos foram realizados a fim de tratar imagens degradadas por meios participativos. Métodos que utilizam a imagem de entrada como única fonte de informação geralmente podem ser classificados de acordo com sua abordagem como métodos de *realce* ou *restauração*.

Os métodos de realce manipulam diretamente algumas propriedades da imagem, *e.g.*, contraste, borramento, e cor, afim de melhorar sua aparência. Alguns exemplos bastante difundidos desta categoria são o CLAHE ([HUMMEL, 1977](#)) e o filtro bilateral ([TOMASI; MANDUCHI, 1998](#)). Fusões dos métodos tradicionais de realce são mencionados na literatura e apresentam resultados bastante satisfatórios, *e.g.*, ([ANCUTI et al., 2012](#); [ANCUTI; ANCUTI, 2013](#); [BAZEILLE et al., 2006](#)).

A manipulação direta das propriedades da imagem, reduz os efeitos da turbidez ao custo de degradar outras informações. Além disso, estas abordagens não costumam considerar a variação espacial que existe em algumas imagens degradadas por meios participativos.

Métodos de restauração se concentram em resolver a equação que descreve a formação da imagem nesses meios, a fim de recuperar a imagem sem turbidez. São considerados métodos de restauração [Fattal \(2008\)](#), [He, Sun e Tang \(2011\)](#), [Berman, Treibitz e Avidan \(2016\)](#), entre outros.

A grande maioria destas abordagens buscam inferir a refletividade dos objetos da cena através da [Equação 2.12](#), ou equações bastantes similares. Para que isso seja possível são necessárias estimativas de transmissão e iluminação. Vale mencionar que muitos destes métodos assumem uma transmissão única para todos canais de cor, o que é bastante problemático especialmente quando o meio participativo não afeta todos os canais de cor igualmente.

Os métodos de restauração mais tradicionais assumem que imagens livres de turbidez possuem algumas propriedades, cujo grau de presença pode indicar diferentes níveis de turbidez em cada região da imagem.

[Fattal \(2008\)](#) assumiu que não há covariância entre a refletância da cena e a iluminação, de modo que a transmissão pode ser definida como a única fonte de covariância. No entanto, [Fattal \(2014\)](#) demonstrou que esta hipótese funciona apenas para condições de baixa degradação.

[Fattal \(2014\)](#) propôs um método para estimar a transmissão que usa o *color line assumption* ([OMER; WERMAN, 2004](#)). Este estima transmissão, por encontrar o ponto de intersecção entre as linhas de cor e o vetor com a orientação da cor do meio. O resultado disto é um método um pouco mais robusto à maiores intensidades de degradação em relação ao ([FATTAL, 2008](#)). Entretanto, este método depende da existência de janelas da imagem que respeitem uma série de condições. Dentre elas estão que (i) os pixels desta janela devem suportar uma linha em \mathbb{R}^3 ; (ii) a janela não deve ser uma borda; (iii) a refletância deve ser positiva quando decomposta usando o seu modelo; (iv) o ângulo da intersecção entre a linha que representa a janela e a fonte de iluminação deve ser suficientemente grande; (v) o ponto de intersecção com a fonte de iluminação deve ser próximo dos pontos de intersecção de outras janelas; (vi) a transmissão estimada deve estar no intervalo $[0, 1]$; e (vii) a iluminação deve variar suficientemente nesta janela. A necessidade de atender esta quantidade de promissas é uma desvantagem desta proposta.

[Tarel e Hautiere \(2009\)](#) usa um balanceamento de branco como pré-processamento para que a contribuição atmosférica passe a ser o branco puro. Sua estimativa de contribuição do meio é uma porcentagem da diferença entre a média local do mínimo entre os canais da imagem e seu desvio padrão. Posteriormente, a estimativa inicial é refinada utilizando uma versão estendida do filtro de mediana.

O algoritmo proposto em ([BERMAN; TREIBITZ; AVIDAN, 2016](#)) assume que imagens livres de névoa podem ser aproximadas de forma satisfatória, usando apenas algumas centenas de cores distintas. As cores da imagem formam grupos com valores bastante próximos no espaço RGB. Os autores observaram também que os grupos frequentemente são formados por pixels dispersos na imagem livre de névoa, e portanto, localizados em diversas partes da cena à diferentes distâncias do observador.

Desta forma, na presença de névoa os grupos passam a formar linhas. O método assume que o ponto mais distante da fonte de iluminação de cada linha não possui névoa. Com base na relação entre cada ponto e o extremo de sua respectiva linha é estimada a transmissão inicial deste ponto. A partir da transmissão inicial é feita uma regularização, onde os pixels são suavizados em relação a estrutura da imagem de entrada e grupos com baixo desvio padrão possuem uma menor influência.

He, Sun e Tang (2011) propôs um método baseado em uma premissa criada pelos autores, conhecida como *Dark Channel Prior*. Nela, o valor mínimo entre canais de cor em uma janela da imagem é usado como um indicativo da transmissão. Esta ideia, apesar de simples, é bastante robusta para névoa. Por este motivo ele deu origem à uma série de métodos que tem por objetivo adaptar esta premissa para ambientes subaquáticos, e.g. (BIANCO; MOHAN; EUSTICE, 2010; CHIANG; CHEN, 2012; GALDRAN et al., 2015; LU et al., 2015; DREWS et al., 2016). No entanto, essas adaptações desconsideram a grande variedade de cores existentes debaixo d'água, assumindo algumas condições específicas como o *Red Channel Absorption* em Galdran et al. (2015).

Meng et al. (2013) propôs um método de restauração para névoa que explora o uso de restrição de limites no mapa de transmissão para estender o He, Sun e Tang (2011). Sua estimativa de transmissão é refinada utilizando um método de otimização baseado em uma normalização na diferença absoluta de transmissão entre pixels vizinhos e em uma função que pondera a influência deles com base nas estruturas da imagem.

Recentemente, foram desenvolvidos métodos onde uma *Convolutional Neural Network* - Rede Neural Convolutiva (CNN) é treinada para estimar a transmissão (CAI et al., 2016a; CAI et al., 2016b). Estes métodos passam por uma etapa de treinamento com diversos pares de imagens e suas respectivas transmissões.

Em Cai et al. (2016a) a CNN estima um valor de transmissão com base em imagens coloridas de 16×16 pixels. Portanto, são apresentados pares de imagens com névoa sintética e um valor único de transmissão para todos pixels e todos os canais. Para realizar a restauração de uma imagem utilizando esse método é necessário refinar a transmissão estimada utilizando He, Sun e Tang (2013). Apenas então, a cor verdadeira é restaurada utilizando a Equação 2.12.

Já em Cai et al. (2016b) a CNN se divide em duas partes, na primeira é feita uma estimativa mais grosseira com base em toda a imagem e na segunda a transmissão é refinada localmente. Durante o treinamento são apresentados imagens de resolução 320×240 pixels e sua respectiva transmissão, onde cada pixel possui um valor próprio de transmissão. Neste método não é feita nenhuma suavização ou refinamento após a estimativa da rede.

Entretanto, em nenhum dos métodos já citados a rede é treinada diretamente para

melhorar o resultado da restauração. Neste contexto, [Li et al. \(2017\)](#) propôs uma rede para *dehazing* que é baseada em um modelo de propagação da luz reformulado. Onde a transmissão e a luz atmosférica são estimadas de forma conjunta e a restauração é otimizada diretamente.

Percebe-se que na revisão apresentada, que entre os métodos envolvendo aprendizado de máquina apenas [Li et al. \(2017\)](#) utiliza uma abordagem fim-a-fim para o processamento de imagens em meios participativos. Porém, vale ressaltar que estes meios causam um efeito não uniforme e dependente da cena, já que a atenuação da componente direta está associada à distância de cada pixel à cena. Com isso é necessário que a rede esteja preparada para lidar com esse tipo de fenômeno.

3 Uma Metodologia para Redução da Turbidez em Imagens Ópticas

Os métodos de restauração que utilizam aprendizado de máquina, como mostrado na [Seção 2.2](#), estimam parâmetros do modelo simplificado de formação da imagem com intuito de recuperar a imagem sem a influência do meio. No trabalho aqui apresentado a estrutura da rede não utiliza o modelo simplificado e, portanto, pode se beneficiar de melhorias na simulação utilizada para gerar o *dataset* sem que a rede precise ser modificada.

A grande maioria das soluções que utilizam *deep learning* para problemas semelhantes realizam a tarefa de reconstrução fim-a-fim, treinando a rede para recuperar a imagem sem degradação com base na imagem degradada ([NAH; KIM; LEE, 2017](#); [LEDIG et al., 2016](#); [ZHANG; ISOLA; EFROS, 2016](#)). Porém, como mencionado anteriormente, o fenômeno tratado neste trabalho varia espacialmente de acordo com a cena capturada fazendo com que *CNNs* percam estruturas da imagem durante o processamento. Para lidar com esse problema propõe-se o uso do *guided filter* ([HE; SUN; TANG, 2013](#)) como parte da rede. Este filtro produz uma saída com base em uma entrada e em uma imagem guia, espera-se que utilizando a imagem de entrada como guia a rede seja capaz manter as estruturas da imagem. O funcionamento deste filtro será detalhado na [Seção 3.1](#).

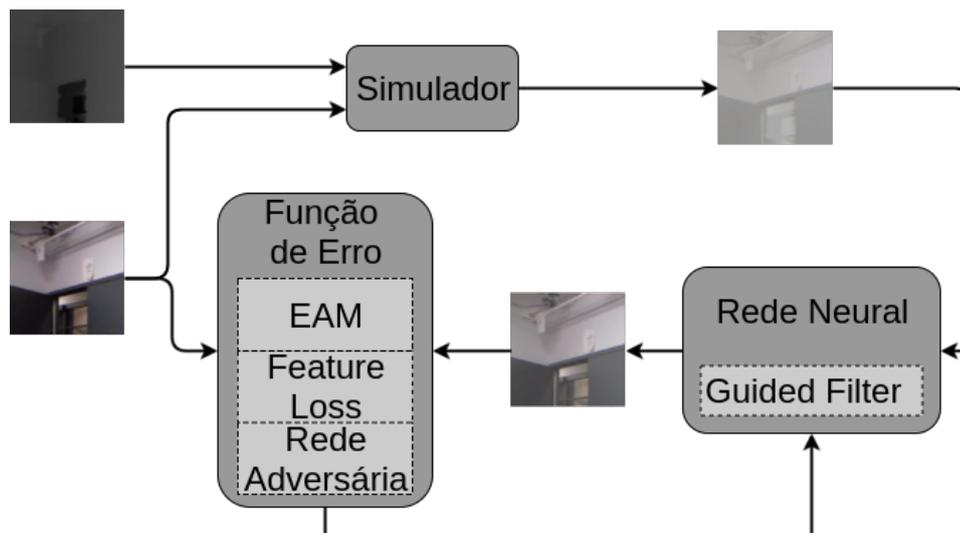


Figura 7: Representação da metodologia utilizada pelo método proposto. Nesta ilustração estão representados o simulador ([Seção 3.4](#)), a função de erro e a rede neural ([Seção 3.2](#)), onde utiliza-se o *guided filter* ([Seção 3.1](#)). Conforme detalhado na [Seção 3.3](#) a função de erro, pode ser uma ou uma combinação das funções representadas pelos retângulos pontilhados.

A [Figura 7](#) apresenta uma visão geral da arquitetura proposta, onde cada um dos módulos serão detalhados neste capítulo. Na [Seção 3.2](#) será mostrada a rede neural proposta para tratar o problema de redução da turbidez em imagens. Na [Seção 3.3](#) é avaliado o uso de diferentes funções de erro na etapa de treinamento desta rede. Por fim, na [Seção 3.4](#) serão discutidos o uso de imagens simuladas e detalhes do simulador utilizado.

3.1 Guided Filter

Os filtros invariantes a translação, como os kernels de uma convolução, possuem pesos constantes e independentes do conteúdo da imagem. Porém, em muitos casos, como no problema de redução da turbidez onde ela varia espacialmente, é interessante o uso de uma imagem guia que incorpora mais informações sobre a estrutura da imagem na filtragem. Muitos trabalhos como por exemplo ([HE; SUN; TANG, 2011](#); [CAI et al., 2016a](#); [MENG et al., 2013](#)), após estimar a transmissão utilizam a imagem em meio participativo como guia para recuperar estruturas perdidas durante a estimativa de transmissão.

Uma das abordagens para aplicar filtros variantes a translação é otimizar uma função quadrática que conforme as informações contidas na imagem guia, força algumas restrições sobre a imagem de saída. Ela é utilizada por [He, Sun e Tang \(2011\)](#) e apesar de produzir bons resultados demanda muito tempo de computação.

Outra abordagem é explicitamente definir um filtro usando a imagem de guia. Em um filtro desse tipo, o valor de um pixel é a soma ponderada dos pixels de uma janela onde os pesos dependem da imagem guia. Conforme descrito pela seguinte equação:

$$q_i = \sum_j W_{ij}(I)p_j, \quad (3.1)$$

onde p é a imagem de entrada, I é a imagem guia, q é a imagem resultante, i e j são índice dos pixels.

O *guided filter* ([HE; SUN; TANG, 2013](#)) define seus pesos baseado na ideia de que a saída do filtro é localmente linear em relação a imagem guia. Assumindo então que q é uma transformação linear de I em uma janela ω_k centrada no pixel k :

$$q_i = a_k I - I_i + b_k, \forall_i \in \omega_k, \quad (3.2)$$

onde a_k e b_k são coeficientes lineares assumidos como constantes em ω_k . Estes coeficientes são determinados através da minimização da diferença entre a imagem de saída e a guia. Isto é feito através da seguinte função de custo:

$$E(a_k, b_k) = \sum_{i \in \omega_k} ((a_k I_i + b_k - p_i)^2 + \epsilon a_k^2). \quad (3.3)$$

Nesta equação ϵ é um parâmetro de regularização que faz com que a_k não assumam valores muito elevados. Após algumas aproximações, os pesos do filtro são definidos como:

$$W_{ij}(I) = \frac{1}{|\omega|^2} \sum_{k:(i,j) \in \omega_k} \left(1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \epsilon} \right), \quad (3.4)$$

onde μ_k e σ_k^2 são a média e a variância de I na janela e $|\omega|$ é o número de pixels em ω_k .

Observando o comportamento da [Equação 3.4](#) é fácil entender como o filtro funciona. Considerando que o filtro está sendo aplicado em uma região correspondente à uma borda em na imagem guia, quando I_i e I_j estão no mesmo lado da borda os termos $I_i - \mu_k$ e $I_j - \mu_k$ possuem o mesmo sinal, quando estão em lados opostos possuem sinais diferentes. Por isso o termo $1 + \frac{(I_i - \mu_k)(I_j - \mu_k)}{\sigma_k^2 + \epsilon}$ é muito menor quando os pixels estão em lados opostos do que no mesmo lado, o que faz com que as bordas sejam preservadas. Além disso, quando aplicado em uma região com pouca variabilidade, ou seja, σ_k^2 muito menor que o parâmetro ϵ , o filtro se aproxima de um passa baixa.

O *guided filter* pode também ser estendido para imagens coloridas. Em casos onde a imagem guiada p possui mais de um canal, o filtro é simplesmente aplicado em todos os canais de forma independente. Porém quando a imagem guia I possui mais de um canal, a [Equação 3.2](#) deve ser reescrita como:

$$q_i = a_k^T I - I_i + b_k, \forall_i \in \omega_k, \quad (3.5)$$

onde I_i é um vetor de cor 3×1 , a_k é um vetor de coeficientes 3×1 , q_i e b_k é um escalar. Logo, o filtro para imagens coloridas se torna:

$$a_k = (\Sigma_k + \epsilon U)^{-1} \left(\frac{1}{|\omega|} \sum_{i \in \omega_k} I_i p_i - \mu_k \bar{p}_k \right), \quad (3.6)$$

$$b_k = \bar{p}_k - a_k^T \mu_k, \quad (3.7)$$

$$q_i = \bar{a}_i^T I_i + \bar{b}_i, \quad (3.8)$$

onde Σ_k é a matriz de covariância 3×3 de I em ω_k , $\bar{a}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} a_k$, $\bar{b}_i = \frac{1}{|\omega|} \sum_{k \in \omega_i} b_k$, $\bar{p}_k = \frac{1}{|\omega|} \sum_{i \in \omega_k} p_i$ e U é uma matriz identidade 3×3 . O uso de uma imagem de guia colorida é interessante pelo fato de algumas bordas bastante aparentes, por vezes se tornarem imperceptíveis em escala de cinza. Nestes casos, o uso da imagem em escala de cinza pode fazer com que algumas destas bordas sejam perdidas durante o processo de filtragem.

3.2 Rede Neural para Redução da Turbidez em Imagens

Uma das técnicas empregadas na rede é o *batch normalization* (IOFFE; SZEGEDY, 2015). Ela é importante para reduzir um efeito conhecido como *internal covariate shift*. Este efeito é responsável por dificultar o treinamento da rede, pois faz com que os pesos das camadas mais profundas da rede precisem se ajustar à diferentes distribuições ao longo do treinamento.

Outra técnica empregada na estrutura da rede é o uso de *skip connections*, ela consiste em concatenar os volumes de ativação de uma camada mais rasa da rede em uma camada mais profunda. Porém, vale ressaltar que o uso desta técnica implica em um aumento considerável na quantidade de parâmetros a serem treinados pela rede.

A estrutura de rede proposta neste trabalho foi escolhida para que cada pixel da saída considere grande parte do contexto da imagem de entrada, a fim de reduzir a ambiguidade. Além disso, a rede proposta tem por objetivo manter algumas estruturas presentes na entrada. Com o intuito de cumprir esses objetivos, a rede ilustrada na Figura 8 possui duas características principais.

A primeira está associada ao fato dela ser fortemente baseada na *unet* (RONNEBERGER; FISCHER; BROX, 2015). Esta decisão foi tomada tendo em vista que uma rede com um formato *encoder-decoder* faz com que o campo receptivo seja bastante grande, mesmo com um número restrito de camadas. Além disso, as *skip connections* presentes nesta arquitetura fazem com que algumas estruturas da entrada sejam mais facilmente transferidas à saída.

A segunda está no uso do *guided filter* para reduzir a perda de informação estrutural, que é um problema bastante comum nesta aplicação, devido ao fato da turbidez ser espacialmente variante. Muitos trabalhos procuram reduzir a perda de estrutural, seja por meio de uma etapa que explicitamente refina sua estimativas de transmissão (HE; SUN; TANG, 2011; FATTAL, 2008; FATTAL, 2014; CAI et al., 2016a), seja na estrutura da rede com o uso de *skip connections* (CAI et al., 2016b; LI et al., 2017).

Essas decisões serão validadas na Seção 4.1, onde o impacto da retirada das estruturas que estão pontilhadas na Figura 8 será testado no conjunto de dados de validação.

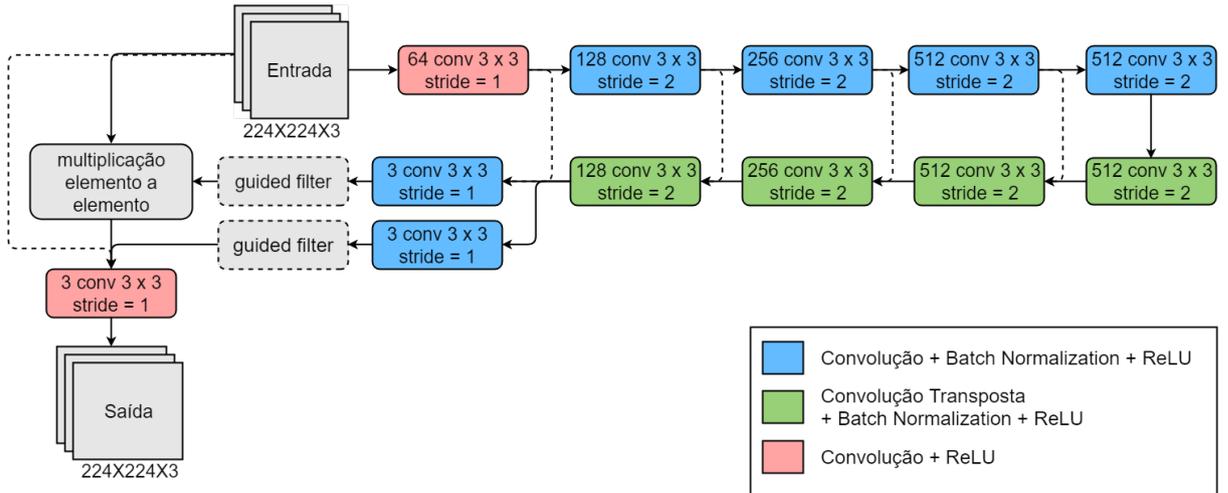


Figura 8: Estrutura de rede utilizada. As camadas representadas em azul e vermelhas são convoluções e as camadas em verde são convoluções transpostas. Todas as camadas convolucionais e convolucionais transpostas fazem convolução com *zero padding* igual a um. As camadas *guided filter* implementam o filtro proposto em He, Sun e Tang (2013), este filtro possui dois parâmetros que não são treinados: o tamanho do filtro que será aplicado e o ϵ . O primeiro valor dentro de cada retângulo que indica convolução ou convolução transposta é o número de *feature maps* de saída. As estruturas que estão pontilhadas terão o seu impacto avaliado na Seção 4.1.

3.3 Funções de Erro

Ao fim da computação de cada amostra apresentada a uma rede neural é preciso realizar o cálculo do erro, para que seja possível então atualizar os pesos da rede de forma a minimizá-lo. No presente trabalho foram avaliadas quatro funções de erro: a *feature loss* (JOHNSON; ALAHI; FEI-FEI, 2016), o Erro Absoluto Médio (EAM), uma combinação de uma função de erro adversária com o EAM e uma combinação de uma função de erro adversária com a *feature loss*.

O EAM procura minimizar a diferença entre a imagem obtida pela rede e a imagem de referência pixel-a-pixel. Esta função foi utilizada pois, apesar de sabidamente não fazer com que a rede aprenda a gerar altas frequências em problemas onde a saída é uma imagem, ela captura de forma bastante satisfatória as baixas frequências. Além disso, gera resultados menos borrados em relação ao erro quadrático médio (ISOLA et al., 2016). A Equação 3.9 apresenta o EAM, onde \hat{y} é a saída obtida pela rede, y é a saída desejada, ambas possuem dimensões $h \times w \times c$, h é a altura, w a largura e c o número de canais.

$$\ell_{EAM} = \frac{\|\hat{y} - y\|_1}{hwc}. \quad (3.9)$$

A *feature loss* (JOHNSON; ALAHI; FEI-FEI, 2016) procura garantir que as imagens de saída e desejada tenham características de mais alto nível similares. Johnson, Alahi e Fei-Fei (2016) usam uma rede neural treinada para classificação, e que portanto

tem em suas primeiras camadas extratores de características semanticamente relevantes. A função de erro é melhor detalhada na [Equação 3.10](#).

$$\ell_{feat}^{\phi_j} = \frac{1}{c_j h_j w_j} \|\phi_j(\hat{y}) - \phi_j(y)\|_2^2. \quad (3.10)$$

Onde $\phi_j(x)$ são as ativações da j -ésima camada da rede utilizada pela função de erro quando processando a imagem x e possui dimensões $h_j \times w_j \times c_j$. Em todos os experimentos feitos neste trabalho, a rede utilizada por esta função de erro é a VGG de 16 camadas ([SIMONYAN; ZISSERMAN, 2014](#)) pré-treinada para o *dataset* ImageNet ([RUSAKOVSKY et al., 2015](#)) e o j foi mantido fixo para a terceira camada convolucional.

A função de erro adversária tem por objetivo aplicar o conceito das *Generative Adversarial Nets* - Redes Adversárias Generativas (GANs) ([GOODFELLOW et al., 2014](#)) para que a saída da rede tenha características semelhantes às imagens sem névoa presentes no conjunto de dados. Para isso, a função de erro consiste em uma rede treinada para discriminar as imagens de saída da rede como falsas e as imagens de referência como verdadeiras. Em paralelo, a rede que gera as imagens com menos turbidez tenta produzir resultados que enganem a rede discriminadora.

Mais especificamente, a rede discriminadora é treinada para produzir um único escalar que representa a probabilidade da imagem recebida por ela ser proveniente do *dataset* e não da saída da rede. Simultaneamente, a rede que produz as imagens é treinada para minimizar a [Equação 3.11](#), isso faz com que ela tente maximizar o erro da rede discriminadora.

$$\ell_{GAN} = -\log D(y), \quad (3.11)$$

onde D representa a rede discriminadora e y é a saída da rede que reduz a turbidez de imagens. A arquitetura utilizada para esta rede, conforme ilustrado na [Figura 9](#), é a mesma proposta em [Ledig et al. \(2016\)](#). Esta rede foi utilizada no presente trabalho visto que o objetivo de uma função de erro para *super resolution* é bastante similar ao do método aqui proposto. Mais especificamente, pretende-se em ambos problemas gerar uma imagem que seja visualmente indistinguível da imagem de referência.

Como a função discriminadora não recebe informação da entrada da rede que produz as imagens, só faz sentido ela ser usada no problema de remoção da turbidez em conjunto com alguma das funções que consideram a entrada ([EAM](#) e *feature loss*). No entanto, como cada uma das funções de erro possui uma faixa de valores distinta é preciso ponderá-las. Com isso as combinações da função de erro adversária com as demais são expressas na [Equação 3.12](#) e na [Equação 3.13](#), onde λ é um parâmetro usado para

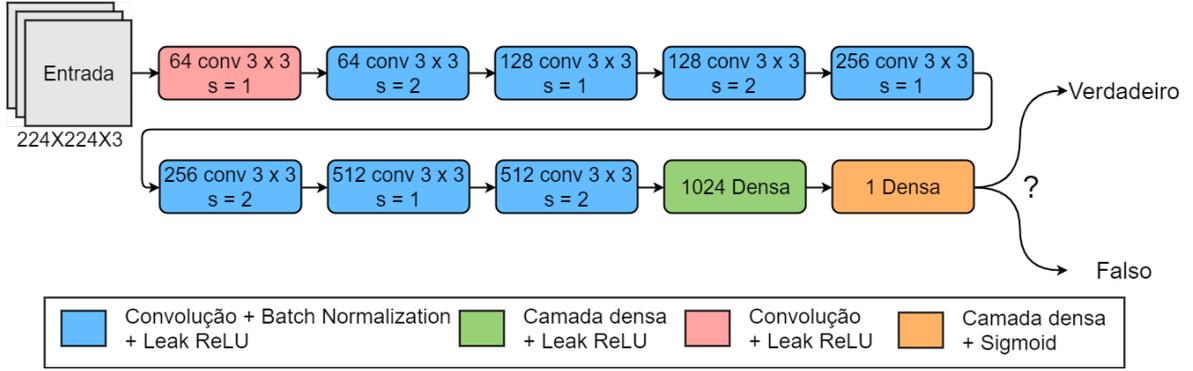


Figura 9: Estrutura utilizada para rede discriminadora. Dentro de cada retângulo que indica convolução e o primeiro valor indica o número de *feature maps* de saída. Nas camadas densas o primeiro valor indica o número de neurônios da camada. Em todas as camadas convolucionais foi usado *zero padding* igual a um.

determinar a influência da função de erro adversária.

$$\ell_{feat+GAN} = \ell_{feat} + \lambda \ell_{GAN}, \quad (3.12)$$

$$\ell_{EAM+GAN} = \ell_{EAM} + \lambda \ell_{GAN}. \quad (3.13)$$

O impacto do uso destas funções de erro no treinamento será avaliado na [Subseção 4.2.1](#). No trabalho aqui apresentado os pesos da rede são ajustados para minimizar a função de erro usando o otimizador *Adam* (KINGMA; BA, 2014). Para evitar o ajuste em demasia para os dados de treinamento, foi separado um conjunto para validação. Nele, o erro é calculado mas não é utilizado para o ajuste dos pesos, de forma que este pode ser usado para avaliar a capacidade de generalização da rede durante o treinamento. Se o erro começar a aumentar para este conjunto significa que o treinamento deve ser encerrado.

3.4 Dados usados no treinamento

O conjunto de dados de treinamento ideal para este problema seria composto por pares de imagens externas com e sem o meio participativo, da mesma cena, com angulação de câmera e condições de iluminação idênticas. Infelizmente, capturar imagens que atendam estas especificações é uma tarefa extremamente difícil, como mostrado em Duarte (2017). Além disso, a quantidade de pares necessária para treinar uma rede neural profunda é bastante grande, o que torna impraticável o uso de um conjunto de dados composto por imagens reais.

Atualmente, simulações físico-realistas de meios participativos, que usam algoritmos de *ray-tracing*, geram imagens bastante semelhantes às reais (PHARR; JAKOB; HUMPHREYS, 2016). Porém, elas possuem um custo computacional bastante elevado

e demandam a modelagem tridimensional das cenas. Dada a quantidade limitada de recursos computacionais deste trabalho, sua limitação de tempo e a quantidade de dados necessária, optou-se por uma simulação mais simplificada tendo como base uma cena real e sua profundidade.

Devido à grande dificuldade associada a captura da distância em cenas externas, as alternativas são usar imagens capturadas em ambientes fechados com o mapa de profundidade ou imagens externas com uma profundidade desconhecida, onde ela pode ser estimada ou constante em toda a imagem. A captura de imagens internas possibilita a simulação de imagens com uma degradação mais consistente com a geometria da cena. Já o uso de imagens externas permite a criação de um conjunto de dados com maior diversidade de cenas e com imagens visualmente semelhantes as que são afetadas por meios participativos. Neste trabalho, devido ao grande impacto da distância entre a câmera e a cena na imagem resultante, optou-se pelo uso de imagens internas.

Duarte (2017) propôs um simulador que, com base em uma imagem sem interferência do meio, seu mapa de profundidade e parâmetros do meio, como coeficiente de espalhamento, simula uma imagem adquirida em um meio participativo da mesma cena. As imagens e o mapa de profundidade usados como entrada do simulador são provenientes do *dataset* proposto em (SILBERMAN DEREK HOIEM; FERGUS, 2012). Vale ressaltar que este *dataset* foi capturado usando o sensor Kinect, e que os mapas de profundidade passaram por uma etapa de pré-processamento onde regiões em que a informação de distância não está disponível foram preenchidas usando a técnica proposta em Levin, Lischinski e Weiss (2004). A Figura 10 mostra algumas imagens sintetizadas usando o simulador para gerar imagens de névoa. Estas imagens são exemplos do *dataset* usado para treinar uma rede capaz de restaurar imagens com névoa no Capítulo 4. Nelas é possível perceber que a informação de distância capturada pelo sensor possui uma série de imperfeições e que as bordas da imagem RGB e de profundidade nem sempre coincidem.

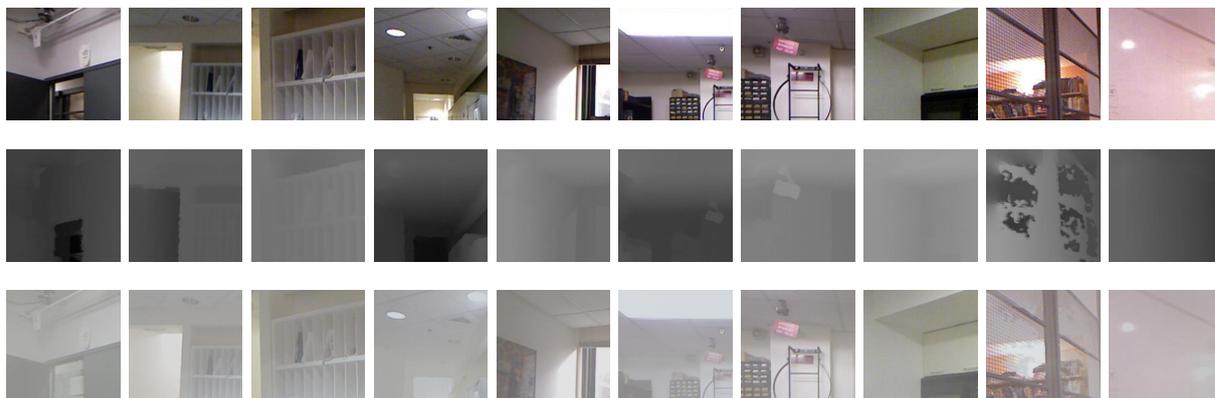


Figura 10: Exemplos de imagens utilizadas no simulador para gerar o conjunto de treinamento. Primeira linha: imagens sem interferência do meio, que correspondem a saída desejada da rede. Segunda linha: Mapas de profundidade usados para simular a degradação da imagem. Terceira linha: imagens com turbidez gerada através de simulação.

4 Resultados Experimentais

Devido à natureza do problema abordado neste trabalho, a avaliação dos resultados obtidos envolve uma série de desafios. Sempre que um método produz como saída uma imagem, o resultado pode ser avaliado de diversas formas. Uma delas é usar modelos matemáticos capazes de estimar a qualidade de um resultado de forma automática. Outra abordagem é utilizar avaliadores humanos para julgar o resultado. Estas duas estratégias possuem desvantagens e não existe um consenso sobre qual delas deve ser adotada para o problema aqui abordado. Portanto, neste trabalho optou-se por avaliar os resultados de forma quantitativa (Seção 4.2), utilizando métricas, e de forma qualitativa, por meio de comparação visual (Seção 4.3).

Existe uma série de métricas para avaliar a qualidade de uma imagem. Elas podem ser divididas em categorias de acordo com a disponibilidade de uma imagem de referência. *Full-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Completa (FR-IQA) onde a imagem de referência com qualidade perfeita está disponível. Assim, a qualidade pode ser mensurada por meio de uma comparação com a imagem referência. *Reduced-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem com Referência Reduzida (RR-IQA), refere-se a categoria onde a imagem de referência não está totalmente disponível. Ao invés disso, apenas algumas características da imagem de referência estão disponíveis e são utilizadas como informação adicional para avaliar a qualidade da imagem testada. No caso da *No-Reference Image Quality Assessment* - Avaliação da Qualidade de Imagem sem Imagem de Referência (NR-IQA), não existe uma referência disponível.

Neste trabalho, a imagem de referência está disponível apenas através do uso de imagens sintéticas, ou seja, FR-IQAs podem ser utilizadas somente para estas imagens. Também, para este problema não existem RR-IQAs e NR-IQAs bem estabelecidas. Neste contexto, o uso de métricas fica limitado à conjuntos de imagens simuladas.

Outra questão pertinente ao uso de métricas de avaliação está na escolha de qual será utilizada, e com isso o critério que ela avalia. O uso de um critério pode favorecer um determinado grupo de métodos. Por esta razão é interessante o uso de métricas, que considerem critérios diferentes. Além disso, é igualmente relevante que as imagens utilizadas possuam características distintas.

Avaliar visualmente os resultados produzidos por métodos de restauração de imagens adquiridas em meios participativos é uma tarefa bastante difícil. Além desta abordagem consumir muito mais tempo em relação ao uso de modelos matemáticos, normalmente o indivíduo que avalia os resultados não segue um critério objetivo. Como um fator compli-

Tabela 1: Parâmetros utilizados no simulador para névoa. Cada um dos parâmetros possui valores diferentes para cada um dos canais de cor.

c			A		
Vermelho	Verde	Azul	Vermelho	Verde	Azul
0.3705	0.3408	0.3214	0.8945	0.9139	0.9268
0.2353	0.2353	0.235	0.9932	0.9932	0.9932
0.4287	0.425	0.4142	0.7994	0.8038	0.8095
0.2787	0.2787	0.2787	0.939	0.939	0.939

gador, o fenômeno tratado varia espacialmente e é bastante frequente que algumas regiões estejam com uma aparência ruim, enquanto outras sejam mais visualmente atrativas. Entretanto, este tipo de avaliação não exige uma imagem de referência e por isso é muito utilizada na literatura.

Em todos os testes que envolvem a rede proposta neste trabalho, ela foi treinada por 25 épocas utilizando o otimizador *Adam* (KINGMA; BA, 2014), com uma taxa de aprendizado 1×10^{-4} . Em cada iteração foram utilizados mini-lotes de 16 imagens de 224×224 pixels, recortadas do *dataset NYU-Depth V2* (SILBERMAN DEREK HOIEM; FERGUS, 2012). Os valores utilizados para a taxa de decaimento exponencial do primeiro (β_1) e segundo (β_2) momentos foram de 0.9 e 0.999, respectivamente. O valor utilizado na constante para estabilidade numérica (ϵ) foi 10^{-8} . O método proposto será comparado com o CLAHE (ZUIDERVELD, 1994), Tarel (TAREL; HAUTIERE, 2009), Ancuti & Ancuti (ANCUTI; ANCUTI, 2013), He et al. (HE; SUN; TANG, 2011), Meng et al. (MENG et al., 2013), Fattal (FATTAL, 2014), MSCNN (CAI et al., 2016b), Dehazenet (CAI et al., 2016a), AOD (LI et al., 2017), *Galdran et al.* (GALDRAN et al., 2015), *Drews et al.* (DREWS et al., 2013) e *Ancuti et al.* (ANCUTI et al., 2012).

Para todos os testes apresentados neste capítulo, as imagens utilizadas na etapa de treinamento da rede foram simuladas utilizando quatro pares de coeficientes de espalhamento (c) e informação proveniente do *backscattering* (A), de acordo com a Tabela 1 para névoa e com a Tabela 2 para água. Cada imagem proveniente do *dataset NYU-Depth V2* é usada para gerar quatro imagens com turbidez, uma para cada par de parâmetros. O resultado da simulação utilizando os parâmetros da Tabela 1 e da Tabela 2 são ilustrados na Figura 11.

4.1 Avaliação das Decisões Tomadas em Relação à Estrutura da Rede

Para avaliar o impacto do uso do *guided filter* durante o treinamento foram utilizadas duas redes, a Dehazenet (CAI et al., 2016a) e a MSCNN (CAI et al., 2016b), que estimam transmissão com base em uma imagem com névoa para posteriormente utilizá-la

Tabela 2: Parâmetros utilizados no simulador para água. Cada um dos parâmetros possui valores diferentes para cada um dos canais de cor.

c			A		
Vermelho	Verde	Azul	Vermelho	Verde	Azul
1.0849	0.1973	1.1536	0.2867	1	0.2341
1.2239	0.1853	1.1875	0.2099	1	0.2349
1.9252	0.1879	1.5317	0.0092	1	0.0567
1.5921	0.0996	1.3438	0.0526	1	0.1402



(a) Entrada



(b)



(c)



(d)



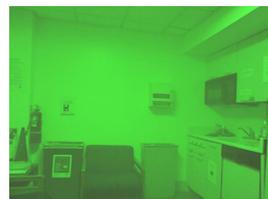
(e)



(f)



(g)



(h)



(i)

Figura 11: Exemplos de simulação utilizando parâmetros para névoa e para água. (a) é a imagem utilizada como base para a simulação. A segunda linha corresponde à simulação utilizando os parâmetros mostrados na Tabela 1, onde (b), (c), (d) e (e) correspondem à terceira, quarta, quinta e última linhas da tabela, respectivamente. A segunda linha desta figura corresponde à simulação utilizando os parâmetros mostrados na Tabela 2, onde as imagens (f), (g), (h) e (i) correspondem à terceira, quarta, quinta e última linhas da tabela, respectivamente.

na restauração. A Dehazenet prevê o uso do *guided filter* como parte de seu pipeline de restauração, porém não o utiliza durante a etapa de treinamento. Enquanto, a MSCNN procura reduzir a perda estrutural em sua estimativa de transmissão por meio do uso de uma *skip connection* em sua arquitetura. Cada uma delas foi comparada com sua respectiva versão que utiliza o *guided filter* em sua saída durante o treinamento, onde a imagem guia é a imagem de entrada.

Os modelos foram treinados em um conjunto de pares de imagem e transmissão, com dimensões $224 \times 224 \times 3$ pixels e 224×224 pixels respectivamente, onde 90% dos dados foram utilizados para o treinamento e 10% para validação. A função de erro utilizada tanto para o treinamento quanto para a validação foi o EAM. A forma com que os modelos foram treinados foi adaptada para que ambos fossem treinados da mesma maneira e, com isso, a comparação fosse mais simples.

Para este teste as redes foram treinadas por 75 épocas utilizando o otimizador *Adam* (KINGMA; BA, 2014), com uma taxa de aprendizado 10^{-5} . Em cada iteração foram utilizados mini-lotes de 16 imagens. Os valores utilizados para a taxa de decaimento exponencial do primeiro (β_1) e segundo (β_2) momento foi de 0.9 e 0.999, respectivamente. A constante para estabilidade numérica (ϵ) utilizada foi de 10^{-8} .

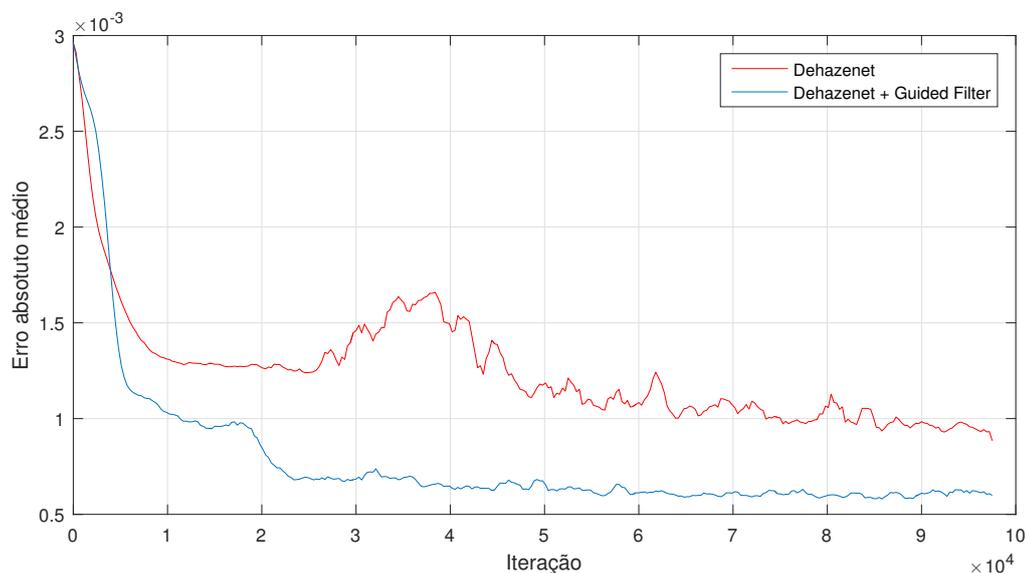


Figura 12: Comparativo entre a Dehazenet (CAI et al., 2016a) e sua versão que utiliza o *guided filter* (HE; SUN; TANG, 2013) durante o treinamento. O valor mostrado é o erro absoluto médio entre a saída do modelo e a transmissão de referência para todo o conjunto de validação. Neste experimentos foram usados $\epsilon = 10^{-3}$ e uma janela de tamanho 40×40 para o *guided filter*.

Como pode ser observado na Figura 12 e na Figura 13, ambas as redes obtiveram um melhor desempenho durante o treinamento que utiliza o *guided filter* em sua saída para o conjunto de validação. Estes resultados indicam que o uso deste filtro durante o treinamento pode ser vantajoso, mesmo em métodos onde a rede procura manter a

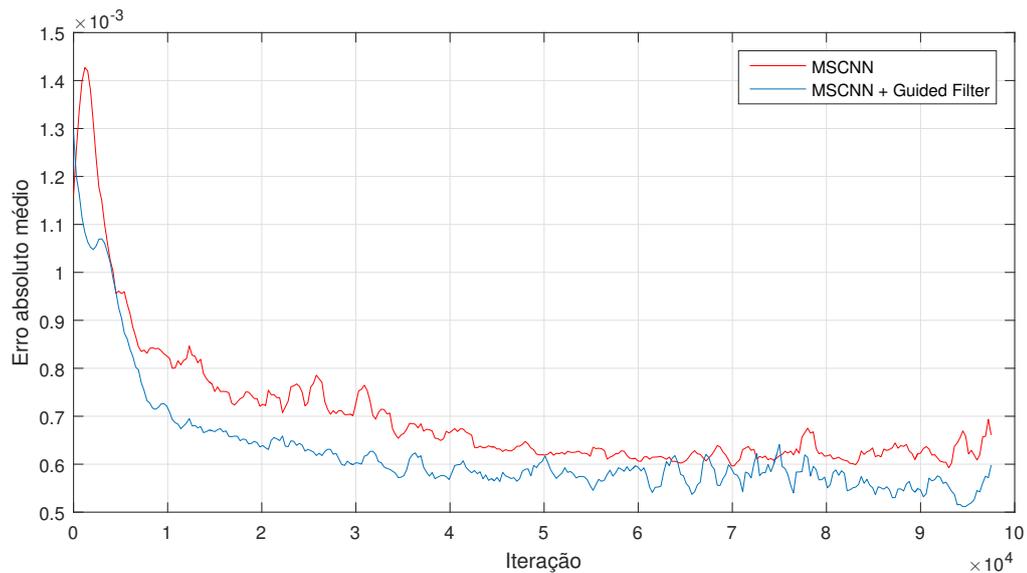


Figura 13: Comparativo entre a MSCNN (CAI et al., 2016b) e sua versão que utiliza o *guided filter* (HE; SUN; TANG, 2013) durante o treinamento. O valor mostrado é o erro absoluto médio entre a saída do modelo e a transmissão de referência para todo o conjunto de validação. Neste experimentos foram usados $\epsilon = 10^{-3}$ e uma janela de tamanho 40×40 para o *guided filter*.

estrutura da imagem através do uso de *skip connections*, como acontece na MSCNN (CAI et al., 2016b).

No presente trabalho, além da utilização do *guided filter*, a rede também conta com *skip connections* para preservar as estruturas da imagem. Para avaliar o impacto destas escolhas nos resultados, foram testadas outras duas versões da rede proposta, uma delas sem *skip connections* e sem *guided filter* e outra com *skip connections* e sem o *guided filter*. Como pode ser observado na Figura 14, o uso de cada uma destas técnicas diminui o erro do método no conjunto de validação. Isto indica que estas escolhas de fato possuem um impacto positivo nos resultados obtidos pelo método.

4.2 Análise Quantitativa Utilizando Imagens Internas Sintéticas

Para análise quantitativa do método proposto foram utilizados o índice de similaridade estrutural (SSIM) (WANG; SIMONCELLI, 2004) e o CIEDE2000 (SHARMA; WU; DALAL, 2005). O uso destas métricas é interessante, pois cada uma delas avalia uma característica distinta da imagem. O SSIM considera a similaridade entre as estruturas da imagem avaliada e da imagem de referência em escala de cinza, enquanto o CIEDE2000 avalia a diferença de cor entre as imagens. Portanto, os métodos com bom desempenho devem apresentar alto SSIM e baixo CIEDE2000.

O conjunto de dados usado nesta comparação foi o D-Hazy (ANCUTI; ANCUTI;

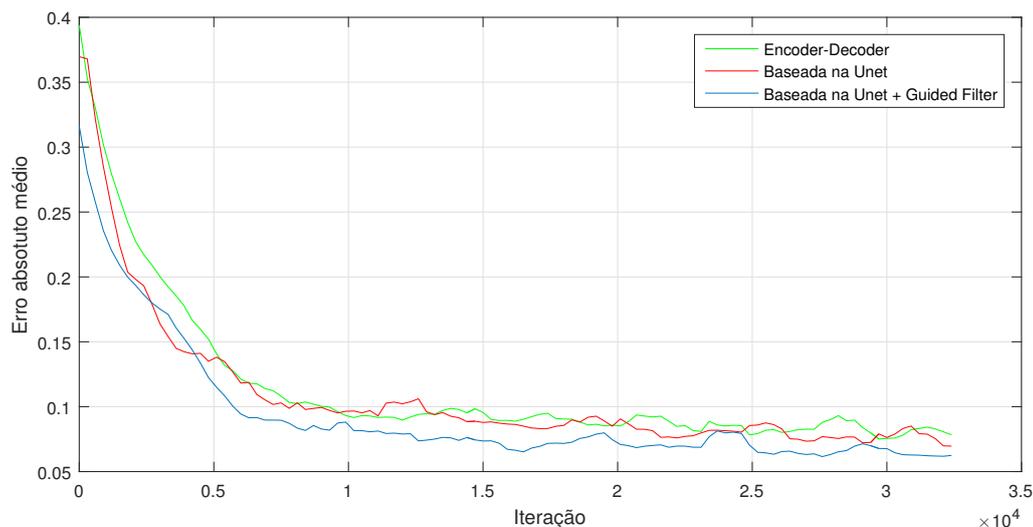


Figura 14: Comparativo entre a rede utilizada neste trabalho (identificada como Baseada na Unet + Guided Filter), sua versão sem *guided filter* (identificada como Baseada na Unet), e sua versão sem *guided filter* e sem *skip connections* (Encoder-Decoder). O valor mostrado é o erro absoluto médio entre a saída do modelo e a saída desejada para todas as imagens do conjunto de validação, durante o treinamento. Neste experimentos foram usados $\epsilon = 10^{-3}$ e uma janela de tamanho 40×40 para o *guided filter*.

VLEESCHOUWER, 2016). Ele é composto de imagens com névoa sintética derivadas dos *datasets Middlebury* (NESIC et al., 2014) e *NYU-Depth V2*. Entretanto, somente a porção proveniente do *dataset Middlebury* (ilustrada na Figura 15) foi utilizada nesta avaliação, já que as imagens derivadas do *dataset NYU-Depth V2* foram utilizadas para gerar as imagens de treinamento e seria uma comparação injusta.

4.2.1 Avaliação do Uso de Diferentes Funções de Erro

Esta subseção é dedicada a avaliar o uso das funções de erro mencionadas na Seção 3.3 utilizando a rede proposta. Os valores de λ utilizados nas funções de erro que combinam o uso de GAN com EAM (Equação 3.13) e com *feature loss* (Equação 3.12) foram de 0.01 e 0.1, respectivamente. Estes parâmetros foram escolhidos por produzirem o melhor resultado dentre os testados. Porém, vale mencionar que por restrições de recursos computacionais e de tempo estes parâmetros não foram exaustivamente testados.

Como pode ser observado na Tabela 3, o impacto das funções de erro nos resultados obtidos pelo método é bastante pequeno neste conjunto de dados. Isto pode ser um indicativo de que o *dataset* utilizado no treinamento não representa de forma satisfatória o fenômeno abordado já que as características penalizadas pelas funções de erro não se refletem no resultado. Por exemplo, seria esperado que a função de erro EAM tivesse um CIEDE2000 mais baixo em relação a *feature loss*, o que não é observado na Tabela 3. Apesar da diferença entre os resultados de redes treinadas com funções de erro distintas

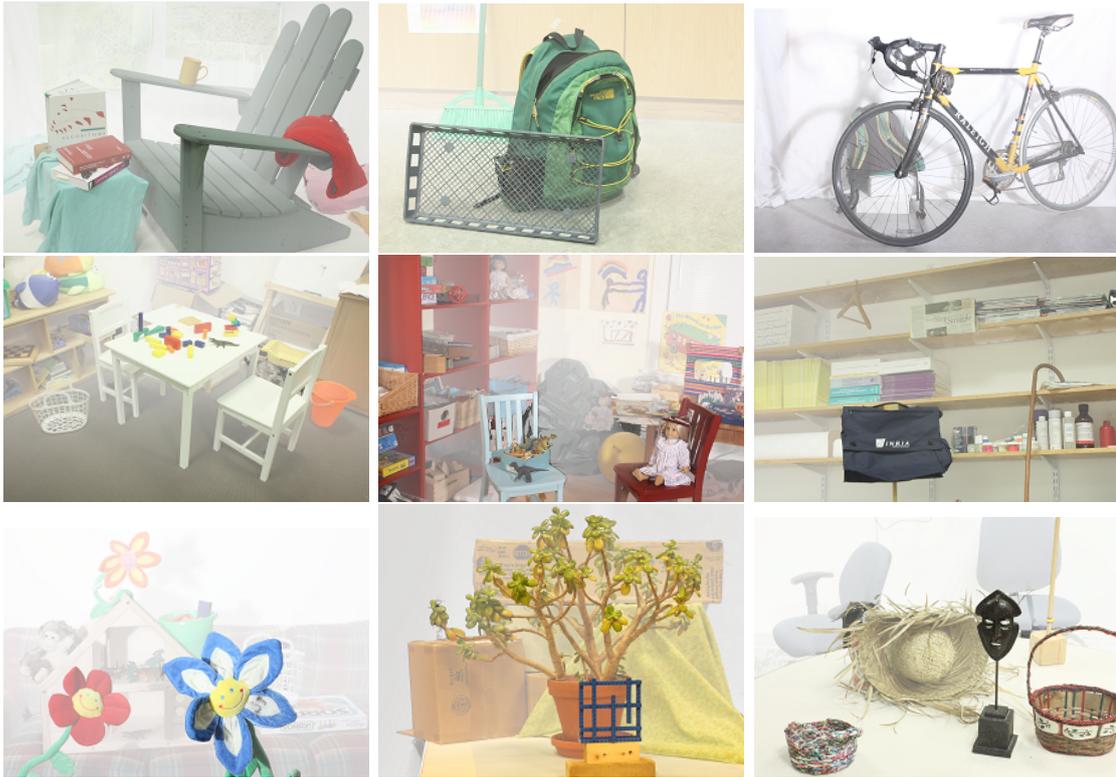


Figura 15: Exemplos de imagens presentes no *dataset* D-Hazy provenientes do *Middlebury*.

não ser muito significativos, o melhor resultado usando a métrica SSIM foi obtido pela *feature loss* enquanto o melhor CIEDE2000 foi obtido pela combinação da *feature loss* com a função de erro adversária.

4.2.2 Avaliação Quantitativa em Relação ao Estado-da-arte

Como pode ser observado na [Tabela 4](#) e na [Tabela 5](#), a abordagem proposta neste trabalho obtém resultados bastante interessantes em relação aos demais métodos, ficando atrás apenas do método descrito em [He, Sun e Tang \(2011\)](#), nas duas métricas, tanto para a *feature loss* quanto para a combinação dela com a função de erro adversária. Porém vale frisar que [He, Sun e Tang \(2011\)](#) possui parâmetros que devem ser ajustados para cada uma das imagens a fim de obter este resultado, enquanto o método aqui proposto utiliza uma rede treinada com os mesmos hiper parâmetros para todas elas. Em relação ao AOD ([LI et al., 2017](#)), que também não possui parâmetros a serem ajustados durante a avaliação, a abordagem adotada mostra resultados superiores nas duas métricas avaliadas.

Tabela 3: Comparativo entre as funções de erro testadas usando imagens da porção do *dataset* D-Hazy derivada do *dataset* Middlebury. Para cada imagem é calculado o SSIM e o CIEDE2000, entre a saída do método e a desejada. Os melhores resultados obtidos para cada imagem, em cada uma das métricas são destacados em negrito e os piores são sublinhados.

	Feature Loss		Feature Loss + GAN		EAM		EAM + GAN	
	SSIM	CIEDE2000	SSIM	CIEDE2000	SSIM	CIEDE2000	SSIM	CIEDE2000
Adirondack	0.88	10.635	0.881	<u>12.42</u>	<u>0.874</u>	11.001	0.877	11.953
Backpack	0.91	8.926	0.917	9.005	0.913	8.242	<u>0.914</u>	<u>9.683</u>
Bicycle1	<u>0.906</u>	<u>10.454</u>	0.911	6.475	0.92	6.867	0.924	8.961
Cable	0.639	25.804	<u>0.619</u>	32.07	0.624	<u>32.746</u>	0.642	26.862
Classroom1	0.803	17.401	0.82	18.181	0.808	18.669	<u>0.772</u>	<u>24.077</u>
Couch	0.843	8.061	0.846	9.649	0.829	9.393	<u>0.737</u>	<u>16.473</u>
Flowers	0.824	11.918	0.809	15.212	0.799	15.591	<u>0.757</u>	<u>19.215</u>
Jadeplant	0.649	22.227	0.635	<u>27.611</u>	0.639	26.204	<u>0.625</u>	26.845
Mask	0.883	10.03	0.886	12.294	<u>0.871</u>	<u>12.952</u>	0.885	12.68
Motorcycle	0.763	13.896	0.814	12.564	0.808	12.511	0.791	<u>14.186</u>
Piano	0.862	5.405	0.894	5.6	0.889	5.751	<u>0.839</u>	<u>9.042</u>
Pipes	<u>0.664</u>	13.05	0.681	14.376	0.682	14.428	0.683	<u>16.707</u>
Playroom	0.848	7.382	0.854	7.783	0.855	7.506	<u>0.835</u>	<u>9.539</u>
Playtable	<u>0.894</u>	7.429	0.916	7.848	0.897	7.661	0.911	<u>8.186</u>
Recycle	<u>0.879</u>	10.507	0.892	9.94	0.908	8.965	0.891	<u>10.761</u>
Shelves	<u>0.887</u>	<u>8.754</u>	0.911	7.964	0.899	7.04	0.89	7.918
Shopvac	0.69	16.853	0.731	21.062	0.72	21.454	<u>0.676</u>	<u>26.37</u>
Sticks	0.935	6.34	0.945	6.167	0.944	6.096	<u>0.93</u>	<u>8.235</u>
Storage	0.857	11.393	0.853	13.883	0.836	14.328	<u>0.817</u>	<u>17.694</u>
Sword1	<u>0.876</u>	10.546	0.883	10.534	0.883	10.448	0.883	<u>11.334</u>
Sword2	<u>0.868</u>	11.26	0.89	12.347	0.879	12.57	0.876	<u>14.362</u>
Umbrella	<u>0.896</u>	10.196	0.901	<u>11.897</u>	0.904	9.964	<u>0.896</u>	10.666
Vintage	<u>0.881</u>	<u>13.114</u>	0.954	6.347	0.919	8.714	0.934	11.795
Média	0.832	11.808	0.845	12.662	0.839	12.57	<u>0.825</u>	<u>14.502</u>

Tabela 4: Avaliação quantitativa do método proposto usando imagens da porção do *dataset* D-Hazy (ANCUTI; ANCUTI; VLEES-CHOUWER, 2016) derivada do *dataset* Middlebury. Para cada imagem é calculado o SSIM entre a saída do método e a desejada. Os valores para o He et al.(HE; SUN; TANG, 2011), Meng et al.(MENG et al., 2013), Fattal(FATTAL, 2014), Ancuti & Ancuti(ANCUTI; ANCUTI, 2013), Tarel(TAREL; HAUTIERE, 2009) e CLAHÉ(HUMMEL, 1977) são aqueles disponibilizados por Ancuti, Ancuti e Vleeschouwer (2016). Já os valores mostrados para MSCNN(CAI et al., 2016b), Dehazenet(CAI et al., 2016a) e AOD(LI et al., 2017) foram calculados utilizando a implementação disponibilizada pelos autores. As últimas quatro colunas correspondem ao método proposto utilizando cada uma das funções de erro testadas. Os melhores resultados obtidos para cada imagem, em cada uma das métricas são destacados em negrito e os piores são sublinhados.

	CLAHÉ	Tarel	Ancuti& Ancuti	He et al.	Meng et al.	Fattal	MSCNN	Dehazenet	AOD	Feature Loss	Feature Loss	EAM	EAM
										+ GAN	+ GAN	+ GAN	+ GAN
Adirondack	<u>0.728</u>	0.851	0.89	0.858	0.882	0.753	0.878	0.601	0.838	0.88	0.881	0.874	0.877
Backpack	<u>0.61</u>	0.874	0.877	0.916	0.89	0.869	0.896	0.775	0.879	0.91	0.917	0.913	0.914
Bicycle1	0.768	0.876	0.899	0.882	0.766	0.809	0.927	<u>0.608</u>	0.924	0.906	0.911	0.92	0.924
Cable	0.499	0.649	0.617	0.71	0.668	0.743	0.607	<u>0.486</u>	0.589	0.639	0.619	0.624	0.642
Classroom1	0.609	0.799	0.886	0.911	0.872	0.87	0.783	0.723	0.742	0.803	0.82	0.808	0.772
Couch	<u>0.613</u>	0.716	0.854	0.908	0.862	0.779	0.667	0.675	0.655	0.843	0.846	0.829	0.737
Flowers	0.703	0.792	0.816	0.876	0.829	0.876	0.736	0.66	0.683	0.824	0.809	0.799	0.757
Jadeplant	0.549	0.687	0.632	0.689	0.699	0.631	0.629	0.58	0.605	0.649	0.635	0.639	0.625
Mask	<u>0.682</u>	0.875	0.86	0.888	0.828	0.886	0.836	0.728	0.861	0.883	0.886	0.871	0.885
Motorcycle	<u>0.761</u>	0.76	0.825	0.816	0.805	0.76	0.785	0.646	0.747	0.763	0.814	0.808	0.791
Piano	0.699	0.79	0.839	0.888	0.855	0.794	0.728	0.624	0.755	0.862	0.894	0.889	0.839
Pipes	0.628	0.678	0.715	0.779	0.751	0.681	0.599	<u>0.482</u>	0.585	0.664	0.681	0.682	0.683
Playroom	0.672	0.792	0.831	0.878	0.822	0.802	0.768	<u>0.558</u>	0.748	0.848	0.854	0.855	0.835
Playtable	0.717	0.831	0.842	0.907	0.865	0.813	0.866	<u>0.688</u>	0.823	0.894	0.916	0.897	0.911
Recycle	0.688	0.895	0.898	0.876	0.871	0.727	0.919	0.673	0.86	0.879	0.892	0.908	0.891
Shelves	<u>0.754</u>	0.883	0.929	0.89	0.889	0.871	0.921	0.906	0.883	0.887	0.911	0.899	0.89
Shopvac	0.669	0.726	0.788	0.827	0.821	0.758	0.62	<u>0.564</u>	0.575	0.69	0.731	0.72	0.676
Sticks	0.715	0.884	0.852	0.948	0.953	0.921	0.95	0.81	0.912	0.935	0.945	0.944	0.93
Storage	<u>0.692</u>	0.853	0.806	0.883	0.87	0.819	0.789	0.764	0.746	0.857	0.853	0.836	0.817
Sword1	<u>0.607</u>	0.862	0.835	0.91	0.892	0.865	0.896	0.707	0.857	0.876	0.883	0.883	0.883
Sword2	<u>0.638</u>	0.872	0.887	0.899	0.849	0.758	0.891	0.758	0.876	0.868	0.89	0.879	0.876
Umbrella	<u>0.574</u>	0.818	0.837	0.838	0.798	0.659	0.884	0.702	0.872	0.896	0.901	0.904	0.896
Vintage	<u>0.711</u>	0.858	0.856	0.917	0.78	0.858	0.959	0.649	0.904	0.881	0.954	0.919	0.934
Média	<u>0.665</u>	0.81	0.829	0.865	0.831	0.796	0.806	0.668	0.779	0.832	0.845	0.839	0.825

Tabela 5: Avaliação quantitativa do método proposto usando imagens da porção do *dataset* D-Hazy (ANCUTI; ANCUTI; VLEES-CHOUWER, 2016) derivada do *dataset* Middlebury. Para cada imagem é calculado o CIEDE2000 entre a saída do método e a desejada. Os valores para o He et al. (HE; SUN; TANG, 2011), Meng et al. (MENG et al., 2013), Fattal (FATTAL, 2014), Ancuti & Ancuti (ANCUTI; ANCUTI, 2013), Tarel (TAREL; HAUTIERE, 2009) e CLAHE (HUMMEL, 1977) são aqueles disponibilizados por Ancuti, Ancuti e Vleeschouwer (2016). Já os valores mostrados para MSCNN (CAI et al., 2016b), Dehazenet (CAI et al., 2016a) e AOD (LI et al., 2017) foram calculados utilizando a implementação disponibilizada pelos autores. As últimas quatro colunas correspondem ao método proposto utilizando cada uma das funções de erro testadas. Os melhores resultados obtidos para cada imagem, em cada uma das métricas são destacados em negrito e os piores são sublinhados.

	CLAHE	Tarel	Ancuti&Ancuti	He et al.	Meng et al.	Fattal	MSCNN	Dehazenet	AOD	Feature Loss	Feature Loss + GAN	EAM	EAM + GAN
Adirondack	11.257	15.195	10.82	10.775	11.106	16.057	12.409	19.03	12.58	10.635	12.42	11.001	11.953
Backpack	14.896	11.721	11.739	10.013	10.629	13.3	9.695	18.194	10.055	8.926	9.005	8.242	9.683
Bicycle1	18.067	8.843	12.669	15.296	23.259	16.448	4.975	20.671	4.475	10.454	6.475	6.867	8.961
Cable	25.054	26.469	24.249	16.379	18.895	13.684	30.437	35.363	32.88	25.804	32.07	32.746	26.862
Classroom1	11.205	23.315	11.839	6.315	9.834	20.742	22.518	25.247	24.329	17.401	18.181	18.669	24.077
Couch	11.519	25.64	13.262	6.736	10.775	22.941	23.151	21.175	21.33	8.061	9.649	9.393	16.473
Flowers	16.91	15.861	14.618	8.646	14.013	8.703	21.279	13.829	24.458	11.918	15.212	15.591	19.215
Jadeplant	26.364	20.187	28.415	19.223	21.418	32.39	29.042	35.068	27.164	22.227	27.611	26.204	26.845
Mask	14.781	11.754	11.603	9.703	14.875	11.823	13.126	15.363	11.616	10.03	12.294	12.952	12.68
Motorcycle	13.026	18.585	14.486	13.706	14.15	17.858	15.436	19.771	16.997	13.896	12.564	12.511	14.186
Piano	9.757	20.301	10.336	6.177	8.861	12.588	17.403	28.387	14.224	5.405	5.6	5.751	9.042
Pipes	16.898	24.285	19.193	9.643	11.541	15.463	25.033	34.343	23.166	13.05	14.376	14.428	16.707
Playroom	11.952	18.367	13.233	7.43	12.07	16.091	15.207	34.503	14.865	7.382	7.783	7.506	9.539
Playtable	9.418	19.06	9.118	8.028	14.119	11.519	12.724	24.553	11.848	7.429	7.848	7.661	8.186
Recycle	13.4	11.89	12.474	14.964	14.846	20.655	7.681	25.572	9.004	10.507	9.94	8.965	10.761
Shelves	7.681	15.202	7.181	11.649	14.413	12.4	6.737	6.758	7.696	8.754	7.964	7.04	7.918
Shopvac	17.659	29.251	19.268	16.287	17.569	32.504	32.392	32.444	35.356	16.853	21.062	21.454	26.37
Sticks	20.073	8.489	15.214	7.213	8.196	10.865	5.611	9.614	6.686	6.34	6.167	6.096	8.235
Storage	13.335	16.636	14.782	10.194	12.086	16.851	18.973	13.215	21.533	11.393	13.883	14.328	17.694
Sword1	ared20.07	12.097	14.809	9.66	10.436	13.822	10.252	22.892	11.277	10.546	10.534	10.448	11.334
Sword2	12.184	13.485	12.051	12.458	15.494	37.965	14.682	21.906	13.765	11.26	12.347	12.57	14.362
Umbrella	20.371	15.13	15.933	20.229	21.489	33.149	12.753	20.097	11.795	10.196	11.897	9.964	10.666
Vintage	14.907	8.236	14.612	10.049	21.796	15.794	7.771	15.747	7.562	13.114	6.347	8.714	11.795
Média	15.252	16.956	14.431	11.338	14.429	18.418	16.056	22.337	16.29	11.808	12.662	12.57	14.502

4.3 Análise Qualitativa Utilizando Imagens Reais Externas

Nesta seção serão mostrados os resultados obtidos pelo método proposto em comparação com os produzidos por métodos estado-da-arte. Para isso foram utilizadas algumas imagens presentes no *benchmark* proposto por Fattal (2014), mais especificamente Figura 16, Figura 17, Figura 19 e Figura 20. Também foi utilizada a Figura 18 introduzida por Cai et al. (2016a), e as imagens subaquáticas que foram inicialmente apresentadas por Ancuti et al. (2012) (Figura 22 e Figura 21).

Na Figura 16 é possível observar que os resultados mais agradáveis visualmente foram obtidos por Fattal (FATTAL, 2014). Porém vale destacar que, além deste método depender do ajuste de parâmetros durante a avaliação, os resultados obtidos por ele erram ao tornar a placa branca no canto superior direito (destacada com o retângulo vermelho) da imagem mais escura. Além dele, os demais métodos também incorrem no mesmo erro. O método aqui proposto, por considerar o contexto daquela região, é capaz de identificar que a placa é realmente branca e não está esbranquiçada por causa da névoa.

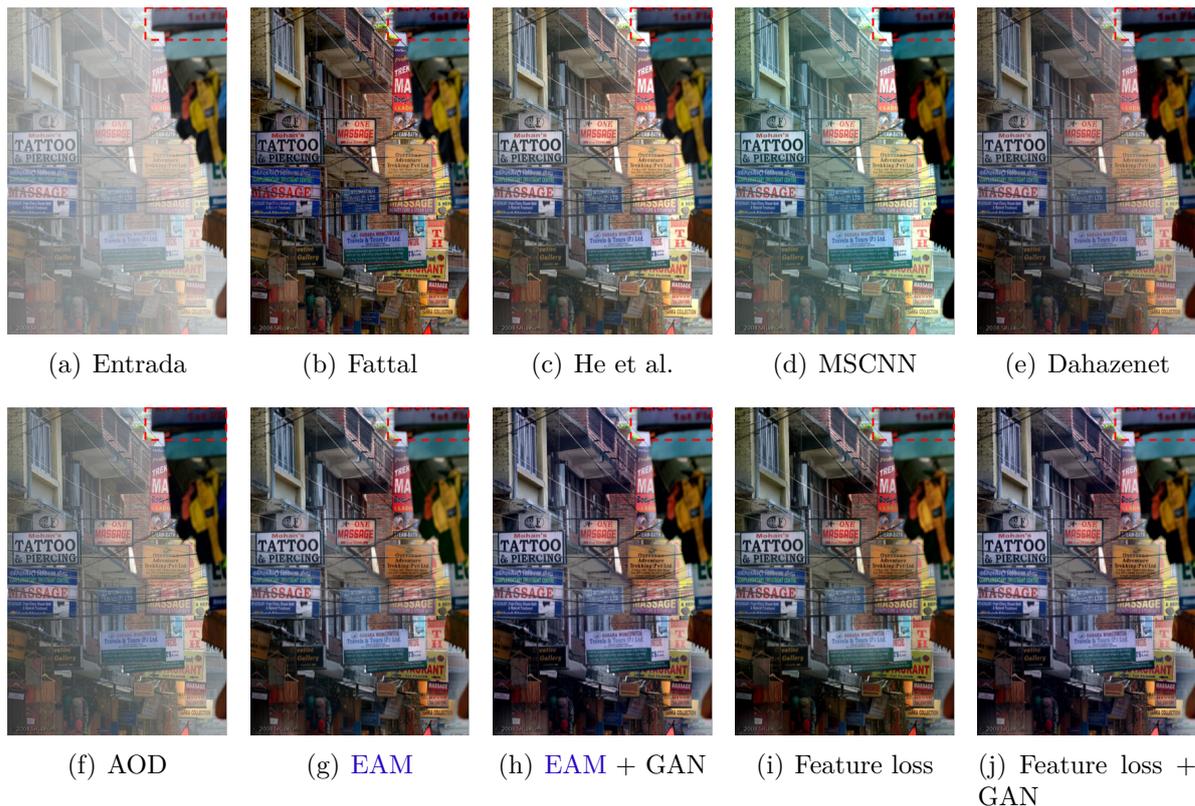


Figura 16: Comparativo utilizando a imagem Logos. Onde (a) é a imagem de entrada, (b), (c), (e) são imagens disponibilizadas pelos respectivos autores, (d) e (f) são imagens geradas pelo código disponibilizado pelos respectivos autores e as demais são imagens geradas pelo método proposto utilizando diferentes funções.

A Figura 17 mostra um exemplo de imagem onde os métodos de restauração de imagens capturadas em meios participativos melhoram a visibilidade. Como pode ser

observado, o método proposto neste trabalho tende a errar em regiões onde há árvores. Isto pode ser causado pelo fato da rede ser treinada com imagens internas onde não existem muitas estruturas parecidas. Mesmo assim é possível notar um ganho de visibilidade significativo em relação à imagem de entrada.

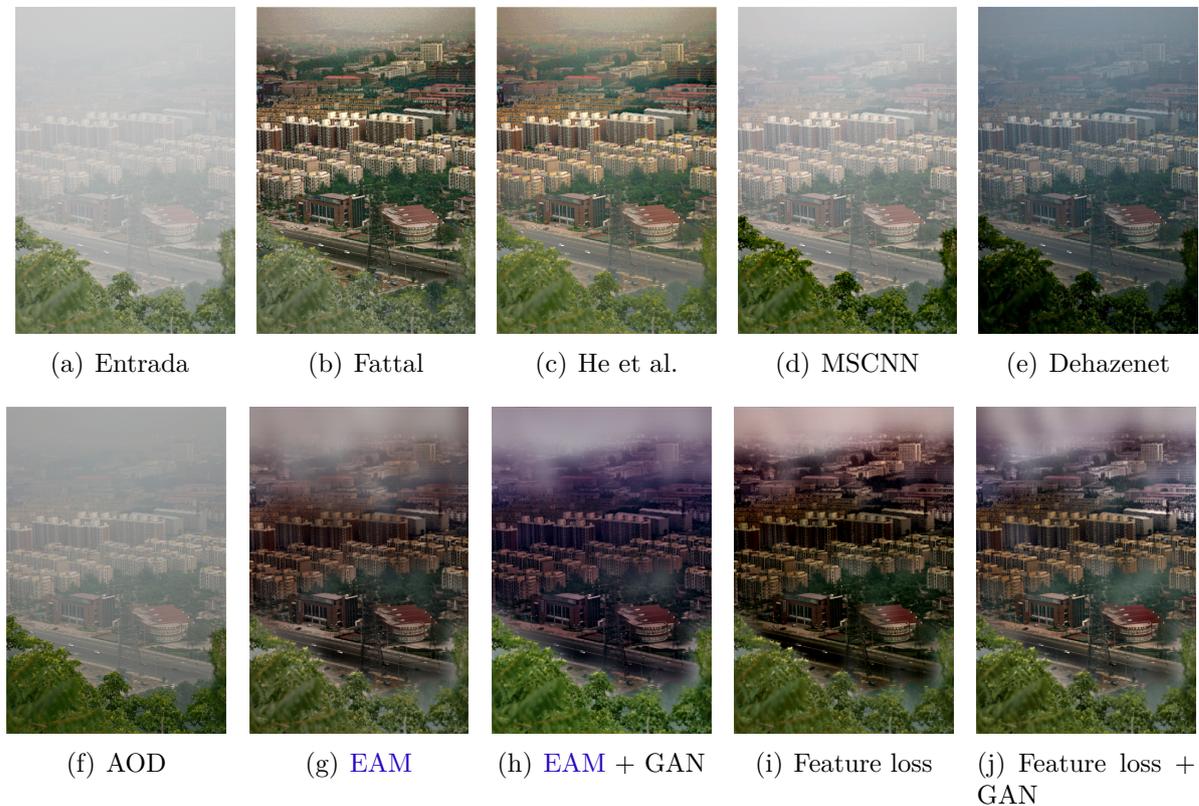


Figura 17: Comparativo utilizando a imagem Cityscape. Onde (a) é a imagem de entrada, (b), (c), (e) são imagens disponibilizadas pelos respectivos autores, (d) e (f) são imagens geradas pelo código disponibilizado pelos respectivos autores e as demais são imagens geradas pelo método proposto utilizando diferentes funções de erro durante o treinamento.

A Figura 18 mostra um caso onde os métodos podem ser aplicados para reduzir o efeito halo (CAI et al., 2016a), que é um arco luminoso em torno da fonte de iluminação. Como pode ser observado o método proposto produz resultados bastante satisfatórios.

A Figura 19 é interessante por apresentar uma mesma estrutura em diferentes distâncias, acarretando diferentes níveis de degradação que devem ser considerados pelos métodos aplicados. Para esta imagem é possível observar que o método gera imagens bastante semelhantes para as diferentes funções de erro.

Na Figura 20 pode-se observar que o método proposto em Fattal (2014) obtém resultados superiores. Porém, é possível também notar a capacidade do método proposto em manter as estruturas da imagem de entrada e sua robustez a uma névoa que não é uniforme em termos de comprimento de onda, ou seja, apresenta diferentes degradações para cada canal de cor.

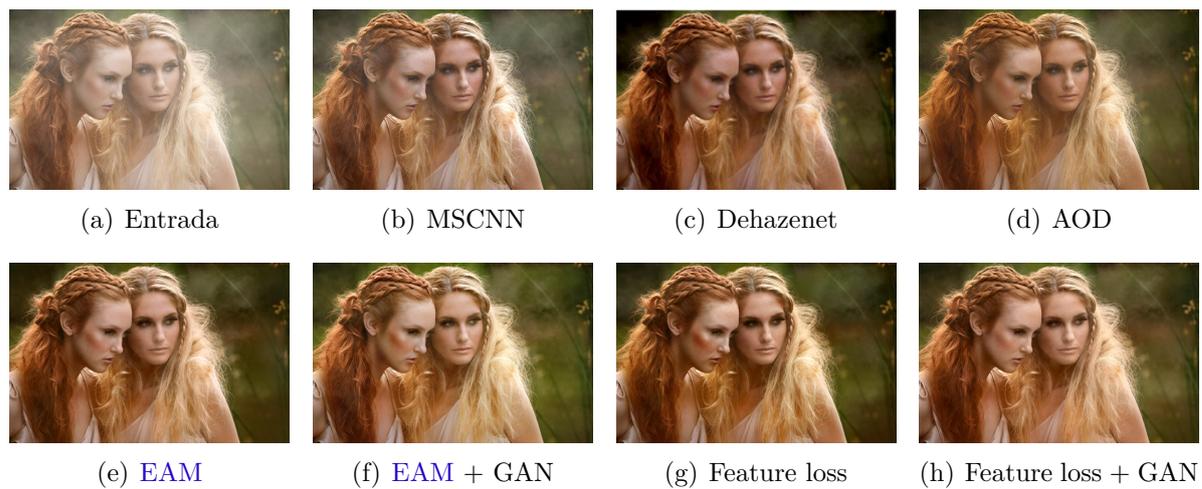


Figura 18: Comparativo utilizando a imagem Girls. Onde (a) é a imagem de entrada, (c) é uma imagens disponibilizadas pelos autores do método, (b) e (d) são imagens geradas pelo código disponibilizado pelos respectivos autores e as demais são imagens geradas pelo método proposto utilizando diferentes funções de erro durante o treinamento.

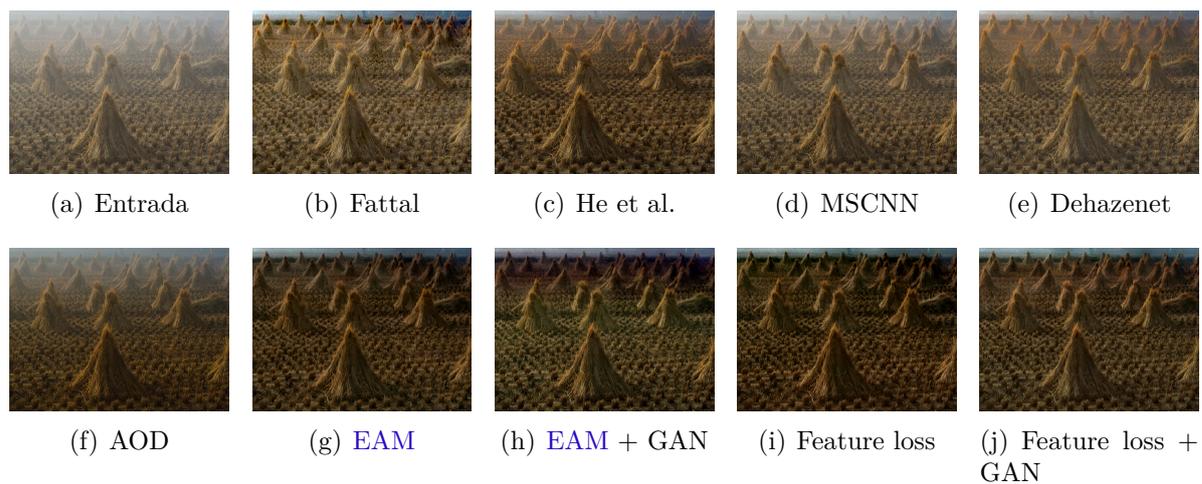


Figura 19: Comparativo utilizando a imagem Wheat Field. Onde (a) é a imagem de entrada, (b), (c), (e) são imagens disponibilizadas pelos respectivos autores, (d) e (f) são imagens geradas pelo código disponibilizado pelos respectivos autores e as demais são imagens geradas pelo método proposto utilizando diferentes funções de erro durante o treinamento.

Na [Figura 21](#) e na [Figura 22](#), é possível observar que o método proposto apesar de fazer uma correção de cor não melhora a visibilidade nestas imagens. É possível que o método não seja capaz de realizar essa correção devido à grande diferença entre as características dos objetos encontrados nestes ambientes e os presentes no conjunto de dados usado durante o treinamento.

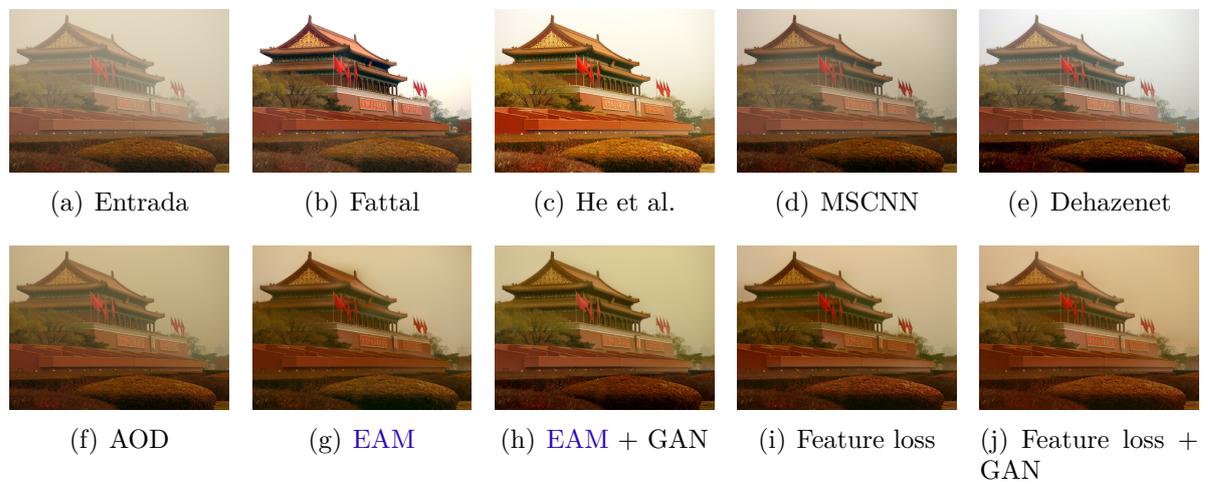


Figura 20: Comparativo utilizando a imagem Tiananmen. Onde (a) é a imagem de entrada, (b), (c), (e) são imagens disponibilizadas pelos respectivos autores, (d) e (f) são imagens geradas pelo código disponibilizado pelos respectivos autores e as demais são imagens geradas pelo método proposto utilizando diferentes funções de erro durante o treinamento.

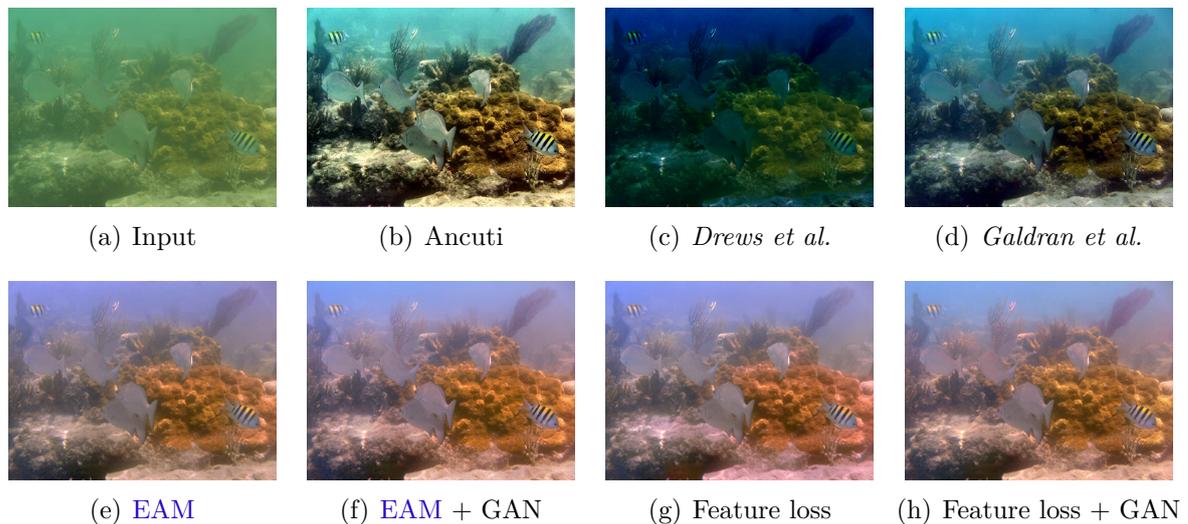


Figura 21: Comparativo utilizando a imagem Peixes. Onde (a) é a imagem de entrada, (b) é uma imagem disponibilizada pelos autores, (c) foi gerada pelo código disponibilizado pelos autores, (d) foi gerada por uma implementação desenvolvida neste trabalho seguindo a descrição dos autores e as demais são imagens geradas pelo método proposto utilizando diferentes funções de erro durante o treinamento.

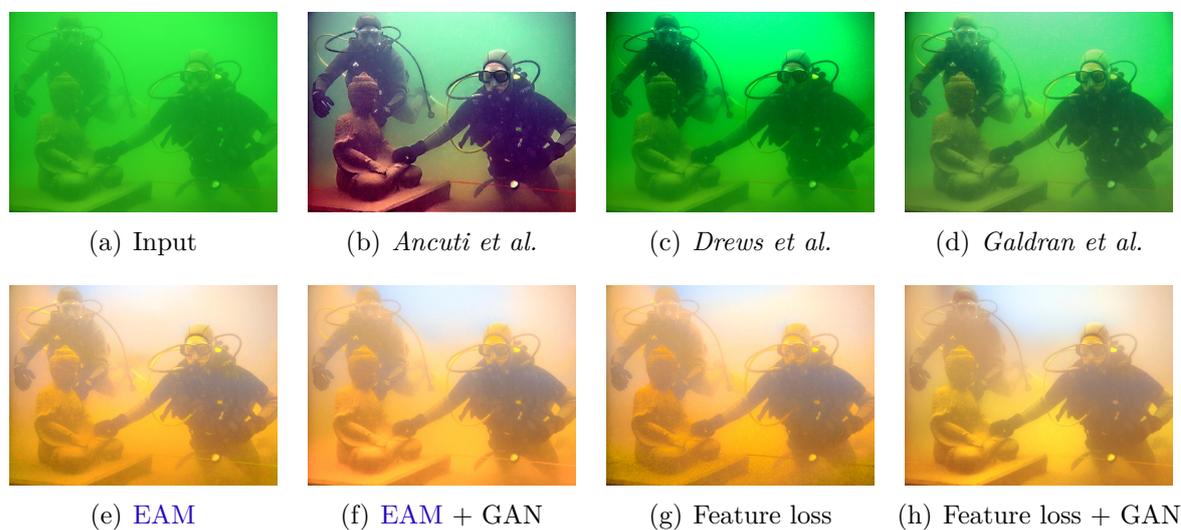


Figura 22: Comparativo utilizando a imagem Mergulhadores. Onde (a) é a imagem de entrada, (b) é uma imagem disponibilizada pelos autores, (c) foi gerada pelo código disponibilizado pelos autores, (d) foi gerada por uma implementação desenvolvida neste trabalho seguindo a descrição dos autores e as demais são imagens geradas pelo método proposto utilizando diferentes funções de erro durante o treinamento.

5 Considerações Finais e Trabalhos Futuros

Esta dissertação apresentou uma metodologia que utiliza redes neurais profundas para reconstrução de imagens degradadas por meios participativos. A abordagem utilizada leva em consideração a dificuldade de manter as estruturas presentes na imagem após o processamento. Através do uso desta abordagem, o método proposto obteve resultados importantes quando comparado à métodos estado-da-arte utilizando as imagens do *dataset* D-Haze. O método apresentou-se superior ao estado-da-arte em técnicas de aprendizado de máquina, bem como mostrou desempenho equivalente aos métodos de reconstrução e realce.

Para o desenvolvimento deste trabalho foram realizados um estudo sobre o fenômeno de propagação da luz em meios participativos e uma revisão dos métodos presentes no estado-da-arte para remoção de turbidez em imagens ópticas. Durante esta revisão observou-se que grande parte métodos possuem uma etapa de refinamento, com o objetivo de reduzir perdas de estruturas das imagens (HE; SUN; TANG, 2011; CAI et al., 2016a; TAREL; HAUTIERE, 2009).

No decorrer do desenvolvimento do método aqui proposto, observou-se que uma grande dificuldade na utilização de CNNs para restauração em imagens é manter a estrutura inicial da imagem que está sendo processada. No entanto, a utilização do *guided filter* (HE; SUN; TANG, 2013) e de *skip connections* como parte da rede reduziu este problema transferindo as estruturas da imagem de entrada para camadas mais profundas da rede.

Durante a avaliação do método foram encontrados indicativos de que o conjunto de dados utilizados para o treinamento da rede é pouco diverso. Pode-se perceber que em imagens com objetos muito diferentes daqueles encontrados no conjunto de treinamento, o método proposto se comporta pior em comparação com os demais métodos. Isso pôde ser observado principalmente em imagens com árvores ou imagens adquiridas em ambientes subaquáticos, onde os objetos possuem cor e textura bem distintas das encontradas em cenas internas.

De acordo com os testes realizados, o uso de diferentes funções de erro impacta muito pouco nos resultados obtidos. Isto pode ser um indicativo de que o conjunto de dados utilizado para o treinamento da rede não captura de forma satisfatória o fenômeno da turbidez.

Uma tarefa fundamental para melhorar os resultados obtidos passa por uma simulação menos simplificada, com maior variabilidade de textura e cor dos objetos. Além disso, uma abordagem interessante seria o uso de ambientes simulados visualmente semelhantes àqueles onde o método será aplicado.

O fato do *guided filter* (HE; SUN; TANG, 2013) ser benéfico para a aplicação explorada neste trabalho levanta algumas questões sobre o desempenho desta técnica em outros contextos. Portanto, pretende-se, futuramente, avaliar o desempenho deste filtro em problemas como colorização e segmentação. Também, pretende-se treinar o parâmetro ϵ em conjunto com os pesos da rede.

Referências

ANCUTI, C. et al. Enhancing underwater images and videos by fusion. *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, 2012. Citado 4 vezes nas páginas 12, 21, 35 e 44.

ANCUTI, C.; ANCUTI, C. O.; VLEESCHOUWER, C. D. D-hazy: A dataset to evaluate quantitatively dehazing algorithms. In: *2016 IEEE ICIP*. [S.l.: s.n.], 2016. p. 2226–2230. Citado 3 vezes nas páginas 39, 42 e 43.

ANCUTI, C. O.; ANCUTI, C. Single image dehazing by multi-scale fusion. *IEEE TIP*, v. 22, n. 8, p. 3271–3282, Aug 2013. ISSN 1057-7149. Citado 4 vezes nas páginas 21, 35, 42 e 43.

ANCUTI, C. O. et al. Locally adaptive color correction for underwater image dehazing and matching. In: *2017 IEEE CVPRw*. [S.l.: s.n.], 2017. p. 997–1005. Citado na página 10.

BAZEILLE, S. et al. Automatic underwater image pre-processing. *Proceedings of the Characterisation du Milieu Marin*, 2006. Citado 2 vezes nas páginas 12 e 21.

BERMAN, D.; TREIBITZ, T.; AVIDAN, S. Non-local image dehazing. In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2016. Citado 3 vezes nas páginas 12, 21 e 22.

BIANCO, N. C.; MOHAN, A.; EUSTICE, R. M. Initial results in underwater single image dehazing. In: IEEE. *OCEANS 2010*. [S.l.], 2010. p. 1–8. Citado na página 23.

CAI, B. et al. Dehazenet: An end-to-end system for single image haze removal. *CoRR*, abs/1601.07661, 2016. Disponível em: <<http://arxiv.org/abs/1601.07661>>. Citado 11 vezes nas páginas 12, 23, 26, 28, 35, 37, 42, 43, 44, 45 e 49.

CAI, Z. et al. A unified multi-scale deep convolutional neural network for fast object detection. In: *ECCV*. [S.l.: s.n.], 2016. Citado 7 vezes nas páginas 12, 23, 28, 35, 38, 42 e 43.

CHEN, L. et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *CoRR*, abs/1606.00915, 2016. Disponível em: <<http://arxiv.org/abs/1606.00915>>. Citado 2 vezes nas páginas 10 e 11.

CHIANG, J. Y.; CHEN, Y.-C. Underwater image enhancement by wavelength compensation and dehazing. *Image Processing, IEEE Transactions on*, IEEE, v. 21, n. 4, p. 1756–1769, 2012. Citado na página 23.

CODEVILLA, F. et al. Achieving turbidity robustness on underwater images local feature detection. *International journal of computer vision*, v. 60, n. 2, p. 91–110, 2015. Citado na página 11.

DREWS, P. et al. Transmission estimation in underwater single images. In: IEEE. *Computer Vision Workshops (ICCVW), 2013 IEEE International Conference on*. [S.l.], 2013. p. 825–830. Citado na página 35.

- DREWS, P. L. et al. Underwater depth estimation and image restoration based on single images. *IEEE Computer Graphics and Applications*, IEEE, v. 36, n. 2, p. 24–35, 2016. Citado na página 23.
- DUARTE, A. C. *Geração de Conjuntos de Dados para Aplicações de Visão Computacional em Ambientes Subaquáticos*. Dissertação (Mestrado) — Universidade Federal do Rio Grande (FURG) - Centro de Ciências Computacionais (C3), 2017. Citado 2 vezes nas páginas 31 e 32.
- FATTAL, R. Single image dehazing. *ACM transactions on graphics (TOG)*, ACM, v. 27, n. 3, p. 72, 2008. Citado 4 vezes nas páginas 12, 21, 22 e 28.
- FATTAL, R. Dehazing using color-lines. *ACM Transactions on Graphics (TOG)*, ACM, v. 34, n. 1, p. 13, 2014. Citado 8 vezes nas páginas 19, 22, 28, 35, 42, 43, 44 e 45.
- GALDRAN, A. et al. Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, Elsevier, v. 26, p. 132–145, 2015. Citado 3 vezes nas páginas 12, 23 e 35.
- GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680. Citado na página 30.
- HE, K.; SUN, J.; TANG, X. Single image haze removal using dark channel prior. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 33, n. 12, p. 2341–2353, 2011. Citado 11 vezes nas páginas 12, 19, 21, 23, 26, 28, 35, 40, 42, 43 e 49.
- HE, K.; SUN, J.; TANG, X. Guided image filtering. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 35, n. 6, p. 1397–1409, 2013. Citado 8 vezes nas páginas 23, 25, 26, 29, 37, 38, 49 e 50.
- HUMMEL, R. Image enhancement by histogram transformation. *Computer graphics and image processing*, Elsevier, v. 6, n. 2, p. 184–195, 1977. Citado 4 vezes nas páginas 12, 21, 42 e 43.
- IOFFE, S.; SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015. Citado na página 28.
- ISOLA, P. et al. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016. Disponível em: <<http://arxiv.org/abs/1611.07004>>. Citado na página 29.
- JAFFE, J. S. Computer modeling and the design of optimal underwater imaging systems. *Oceanic Engineering, IEEE Journal of*, IEEE, v. 15, n. 2, p. 101–111, 1990. Citado 2 vezes nas páginas 15 e 16.
- JASIOBEDZKI, P. et al. Underwater 3d mapping and pose estimation for roV operations. In: IEEE. *OCEANS 2008*. [S.l.], 2008. p. 1–6. Citado na página 12.
- JOHNSON, J.; ALAHI, A.; FEI-FEI, L. Perceptual losses for real-time style transfer and super-resolution. In: SPRINGER. *European Conference on Computer Vision*. [S.l.], 2016. p. 694–711. Citado na página 29.

- KINGMA, D.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. Citado 3 vezes nas páginas 31, 35 e 37.
- LEDIG, C. et al. Photo-realistic single image super-resolution using a generative adversarial network. *CoRR*, abs/1609.04802, 2016. Disponível em: <<http://arxiv.org/abs/1609.04802>>. Citado 3 vezes nas páginas 10, 25 e 30.
- LEVIN, A.; LISCHINSKI, D.; WEISS, Y. Colorization using optimization. In: ACM. *ACM Transactions on Graphics (ToG)*. [S.l.], 2004. v. 23, n. 3, p. 689–694. Citado na página 32.
- LI, B. et al. Aod-net: All-in-one dehazing network. In: *The IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. Citado 7 vezes nas páginas 12, 24, 28, 35, 40, 42 e 43.
- LU, H. et al. Contrast enhancement for images in turbid water. *JOSA A*, Optical Society of America, v. 32, n. 5, p. 886–893, 2015. Citado na página 23.
- MCGLAMERY, B. A computer model for underwater camera systems. In: INTERNATIONAL SOCIETY FOR OPTICS AND PHOTONICS. *Ocean Optics VI*. [S.l.], 1980. p. 221–231. Citado 2 vezes nas páginas 16 e 18.
- MENG, G. et al. Efficient image dehazing with boundary constraint and contextual regularization. In: *Proceedings of the 2013 IEEE ICCV*. Washington, DC, USA: IEEE Computer Society, 2013. p. 617–624. ISBN 978-1-4799-2840-8. Disponível em: <<http://dx.doi.org/10.1109/ICCV.2013.82>>. Citado 6 vezes nas páginas 19, 23, 26, 35, 42 e 43.
- NAH, S.; KIM, T. H.; LEE, K. M. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. [S.l.: s.n.], 2017. Citado na página 25.
- NESIC, N. et al. *Middlebury Dataset*. 2014. Disponível em: <<http://vision.middlebury.edu/stereo/data/scenes2014/>>. Citado na página 39.
- OMER, I.; WERMAN, M. Color lines: Image specific color representation. In: IEEE. *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*. [S.l.], 2004. v. 2, p. II–946. Citado na página 22.
- PHARR, M.; JAKOB, W.; HUMPHREYS, G. *Physically based rendering: From theory to implementation*. [S.l.]: Morgan Kaufmann, 2016. Citado na página 31.
- RONNEBERGER, O.; FISCHER, P.; BROX, T. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. Disponível em: <<http://arxiv.org/abs/1505.04597>>. Citado na página 28.
- ROSENBLUM, M.; GOTHARD, B. A high fidelity multi-sensor scene understanding system for autonomous navigation. In: IEEE. *Intelligent Vehicles Symposium, 2000. IV 2000. Proceedings of the IEEE*. [S.l.], 2000. p. 637–643. Citado na página 12.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, v. 115, n. 3, p. 211–252, Dec 2015. ISSN 1573-1405. Disponível em: <<https://doi.org/10.1007/s11263-015-0816-y>>. Citado na página 30.

- SCHECHNER, Y. Y.; KARPEL, N. Recovery of underwater visibility and structure by polarization analysis. *Oceanic Engineering, IEEE Journal of*, IEEE, v. 30, n. 3, p. 570–587, 2005. Citado 3 vezes nas páginas 12, 17 e 19.
- SHARMA, G.; WU, W.; DALAL, E. N. The ciede2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations. *Color Research & Application*, Wiley Online Library, v. 30, n. 1, p. 21–30, 2005. Citado na página 38.
- SILBERMAN DEREK HOIEM, P. K. N.; FERGUS, R. Indoor segmentation and support inference from rgbd images. In: *ECCV*. [S.l.: s.n.], 2012. Citado 2 vezes nas páginas 32 e 35.
- SIMONYAN, K.; ZISSERMAN, A. Very deep convolutional networks for large-scale image recognition. *CoRR*, abs/1409.1556, 2014. Disponível em: <<http://arxiv.org/abs/1409.1556>>. Citado na página 30.
- SZEGEDY, C. et al. Inception-v4, inception-resnet and the impact of residual connections on learning. *arXiv preprint arXiv:1602.07261*, 2016. Citado na página 11.
- TAREL, J.; HAUTIERE, N. Fast visibility restoration from a single color or gray level image. In: IEEE. *IEEE 12th ICCV*. [S.l.], 2009. p. 2201–2208. Citado 6 vezes nas páginas 12, 22, 35, 42, 43 e 49.
- TOMASI, C.; MANDUCHI, R. Bilateral filtering for gray and color images. *Sixth International Conference on Computer Vision, 1998.*, p. 839–846, 1998. Citado na página 21.
- WANG, H. S. Z.; SIMONCELLI, E. P. Image quality assessment: from error measurement to structural similarity. *IEEE Trans. Image Processing*, IEEE, v. 13, 2004. Citado na página 38.
- ZHANG, R.; ISOLA, P.; EFROS, A. A. Colorful image colorization. In: *ECCV*. [S.l.: s.n.], 2016. Citado na página 25.
- ZUIDERVELD, K. Graphics gems iv. In: HECKBERT, P. S. (Ed.). San Diego, CA, USA: Academic Press Professional, Inc., 1994. cap. Contrast Limited Adaptive Histogram Equalization, p. 474–485. ISBN 0-12-336155-9. Disponível em: <<http://dl.acm.org/citation.cfm?id=180895.180940>>. Citado na página 35.