7 RESULTADOS

Em termos objetivos, a proposição de uma metodologia para melhoria na qualidade de imagens subaquáticas deve estar orientada a redução das dificuldades impostas pelo ambiente à percepção da cena. Os resultados estão sujeitos à condição de problema mal posto que permeia a área. Pois, qual a real percepção das cores, tons e contrastes?

Para Rittel and Webber (1973), as soluções para um problema mal-posto não são do tipo verdadeiras-ou-falsas, mas sim boas-ou-ruins e problemas mal-postos não têm um conjunto enumerável (ou explicável) de possíveis soluções. Estas considerações subsidiam ou explicam a consistente oferta de metodologias para melhoria de imagens UW, em especial para realce, presente na bibliografia.

A real percepção de cores, tons e contraste em uma cena subaquática pode ser alcançada, em uma perspectiva simplificada, quando a cena é exposta à luz solar e sem a presença da água. Entretanto, esta condição não está presente na ampla maiorias das situações práticas, e os resultados dos métodos para melhoria na qualidade sugerem ou usam aproximações. Neste panorama, abordagens que obtém resultados a partir do emprego de imagens de referência sintetizadas (*datasets* pareados) constituem-se em inferências baseadas em inferências. Ao mesmo tempo, esta questão tensiona a crença na efetividade dos critérios quantitativos na avaliação dos resultados. Um modelo treinado sob elevado grau de supervisão (dados pareados), pode apresentar elevado desempenho quando seus resultados são avaliados com métricas que usam imagens de referência geradas pela mesma metodologia daquelas empregadas no treinamento. Entretanto, a adoção deste procedimento é recorrente na bibliografia onde há o emprego de imagens sintetizadas na avaliação de resultados baseada em métricas com referência.

Neste trabalho, o ponto de partida da proposta para melhoria da qualidade é a imagem UW (ou *frame de vídeo*) na condição em que foi adquirida. Assume-se que toda a informação para a efetivação da tarefa de realce deve ser obtida da imagem, sem processamento prévio, ou emprego de versão de referência. Na inferência dos resultados, o método opera sobre o que a imagem original oferece. Assim, esta imagem se constitui na referência efetiva para a tarefa de realce. Parte-se daquilo que se dispõe, do que se conhece, mesmo que, em muitos casos, seja minimamente percebido na observação visual. A percepção da cena é afetada pela água, materiais em suspensão e a respectiva interação destes com a luz ambiente. Este fatores definem a degradação. O método considera a estimativa desta degradação como informação crucial para a sua redução ou remoção na cena realçada.

Nas seções a seguir são apresentados e discutidos os resultados para melhoria de imagem e vídeo, obtidos com a metodologia proposta neste trabalho. Importante destacar o emprego de imagens e vídeos reais, nos quais o efeito de elementos que favorecem a degradação é perceptível e ocorre em diferentes graus de intensidade. Este aspecto também está relacionado com o potencial para aplicação, onde diferentes cenários são esperados, assim como, um desempenho satisfatório do método.

7.1 Resultados para o realce de imagens

As imagens de saída apresentadas nesta seção foram obtidas ao final do treinamento da rede. No método, a imagem de entrada é a única informação requerida, não sendo utilizado pré-processamento.

7.1.1 Imagens de saída do modelo

Na Figura 7.1 é apresentado grupo de imagens associadas à uma amostra do *dataset* UW-Data, correspondendo à entrada, às versões guias de realce e degradação, à imagem degradada pelo MD e a saída da rede. Em conjunto, estão os respectivos histograma e distribuição no espaço de cores.

Percebe-se visualmente que a a imagem-guia de degradação aparenta acentuados turbidez e *color cast*, com perda de cor e contraste. O histograma confirma esta observação, registrando o desequilíbrio entre os canais de cores, com redução significativa nas intensidades e na faixa dinâmica. De forma semelhante, o efeito na distribuição no espaço de cores provoca redução no volume, com o predomínio da cor verde, resultado da intensificação no *color cast*. A imagem-guia para realce mostra a efetividade da expansão do histograma, seguida do reforço na cor, contraste e textura. Há forte redução da turbidez e *color cast*. Embora, o histograma aponte a melhor distribuição dos canais de cores em relação ao histograma da imagem de entrada, ainda persiste a informação residual de *color cast*. Isto é percebido pela predominância do canal verde na região de alta intensidade do histograma. Na imagem, isto pode ser atestado na tonalidade verde do fundo e na parte inferior central da cena.

A imagem de saída do módulo de degradação indicada na Fig. 7.1(d), mostra a os efeitos de ambos os *pipelines* sobre a imagem de entrada. A imagem mostra a intensificação do *color cast* e o *blurring* gerado pela filtragem gaussiana. Os canais de cores encontram-se segregados na intensidade, com o vermelho com menores amplitudes, azul e verde nas faixas moderada e alta, respectivamente. Salienta-se que há informação em toda faixa das intensidades. Isto é importante para a recuperação de cor e contraste durante a decodificação da imagem de saída da rede.

A imagem de saída apresenta melhor definição nas cores na cena, com destaque para os objetos em primeiro plano. A recuperação da textura associada a cor é percebida nos contornos e extremidades dos tentáculos da anêmonas. A informação residual do *color cast* presente na Fig. 7.1(c), foi reduzida, com a região do fundo da cena em coloração aparentando tons naturais. Diferente dos casos anteriores, há equilíbrio na distribuição dos canais de cores.

A contribuição dos módulos dedicados à textura em ambos os *pipelines* pode ser observada nas imagens da Fig. 7.3. São mostradas imagens com diversos elementos associados à textura e os objetos na cena próximos à câmera, facilitando a percepção e comparação visuais. As imagens na Fig. 7.3(a) correspondem as imagens de entrada do modelo. A Fig. 7.3(b) mostra imagens geradas pelo modelo sem a presença dos módulos de textura. As imagens são realçadas, apresentando recuperação de cor e contraste, com redução na turbidez. Entretanto, nota-se a

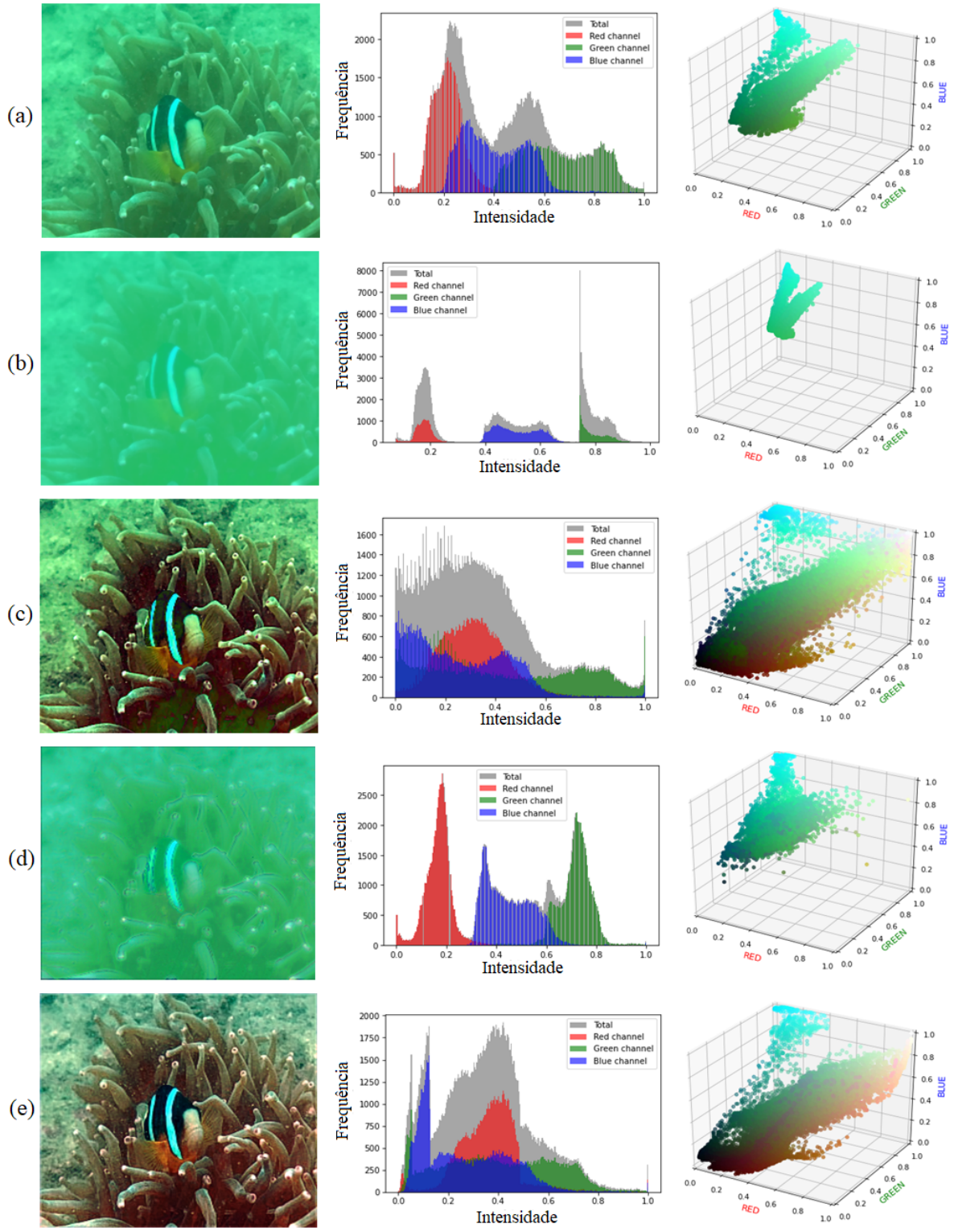


Figura 7.1: Exemplo de imagens resultantes da função de degradação e respectivos histogramas e distribuição no espaço de cores RGB. (a) Imagem de entrada $I_{(k)}$, (b) Versão-guia para degradação $I_{d(k)}$, (c) Versão-guia para realce $I_{e(k)}$, (d) Imagem de saída do MD $\hat{I}^{*}_{(k)}$ e (e) Imagem de saída da rede $\hat{I}_{(k)}$.

presença de *blurring* em bordas e contornos dos objetos. Nas imagens da Fig. 7.3(c) são as obtidas com os módulos para textura presentes em ambos os *pipelines*. Percebe-se a redução no



Figura 7.2: Amostras de imagens de entrada e resultantes com e sem a presenças dos módulos para textura no modelo. (a) Imagem de entrada, (b) Saída sem módulos para textura, (c) Saída com módulos para textura.

borramento, permitindo melhor definição dos elementos de textura na superfície dos objetos. Paralelamente, há melhora no desempenho associado à recuperação de cor, mas, principalmente, de contraste.

A Figura 7.2 apresenta amostras adicionais de imagens de entrada e as respectivas imagens de saída do MD e da rede. As imagens mostram diferentes *color casts* e nível de turbidez moderado, estando os objetos distantes da câmera. O método tem a característica de preservar a cor, mesmo após a redução do *color cast* e turbidez. Este aspecto é importante, pois a única informação disponibilizada ao método é a imagem original. Em geral, há melhoria na percepção das bordas e contornos, resultado da ação sobre a textura no MD. Em regiões da cena em que os objetos encontram-se distantes da câmera, a remoção da turbidez é menos intensa, devido à oclusão parcial gerada pela turbidez. Bordas e contornos já se encontram bastante difusos e a filtragem gaussiana não acrescenta degradação significativa à estas *features*.

Um conjunto de imagens com os objetos próximos à câmera é mostrado na Fig. 7.4. Nestas imagens o efeito da distância é minimizado, com os objetos nas cenas bastante próximos ao dispositivo de captura e a aparente turbidez nas imagens é de moderada a baixa. O detalhamento visual da textura da superfície dos objetos é muito maior em relação à imagem de entrada.

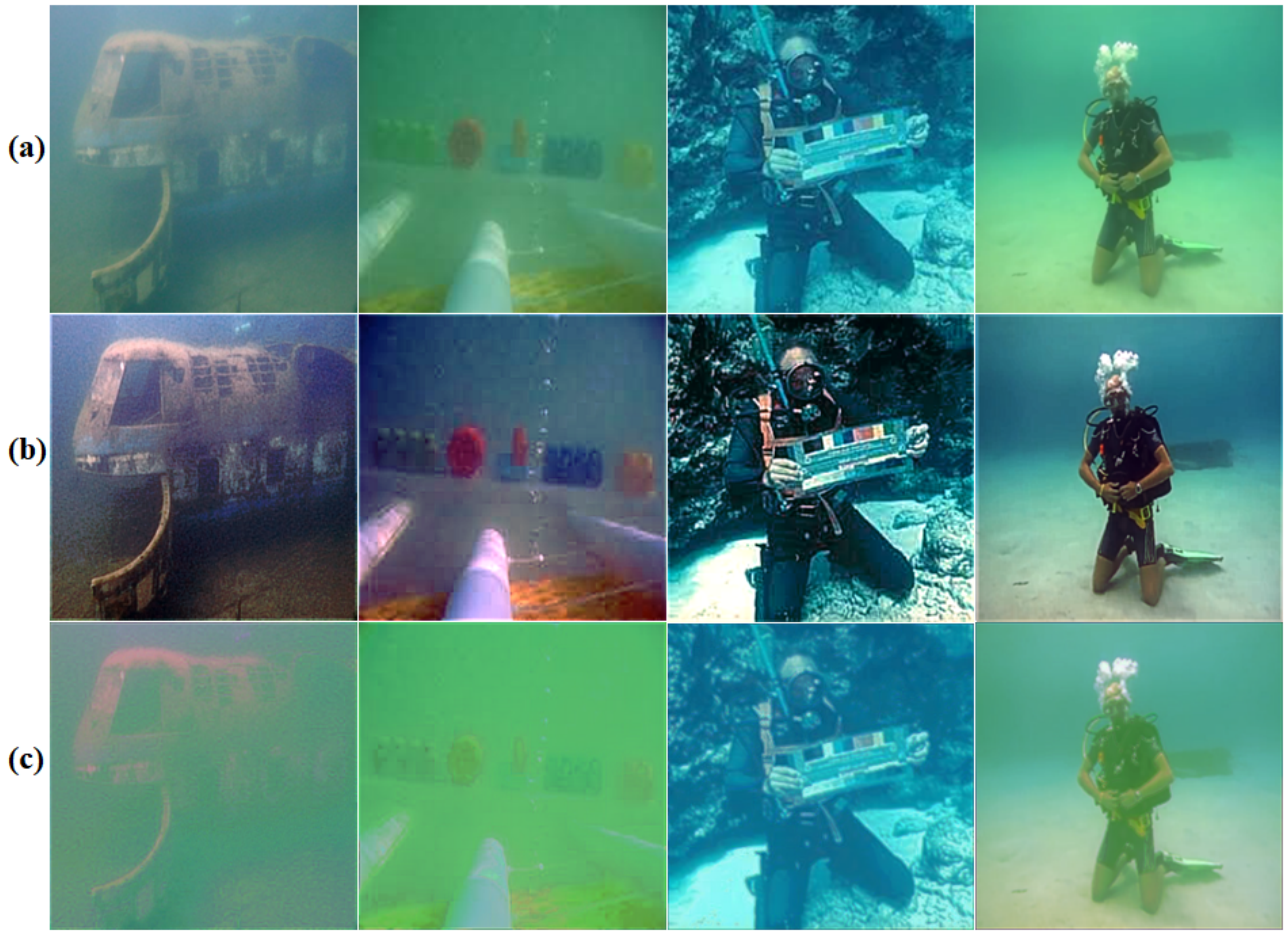


Figura 7.3: Exemplos de imagens de entrada e resultantes da função de degradação e saída da rede. (a) Imagem de entrada $I_{(k)}$, (b) Imagem de saída da rede $\hat{I}_{(k)}$, (c) Imagem de saída do módulo de degradação $\hat{I}^{*}_{(k)}$. Imagens obtidas ao final do treinamento da rede.

Interessante o fato de que a tonalidade predominante na imagem degradada de saída de MD, reproduz a do *color cast* de maneira acentuada, assim como o efeito da filtragem gaussiana na produção do *blurring* sobre bordas e contornos. Estes dois fatores geram penalização na função de perda, conduzindo o aprendizado da rede à sua remoção da imagem realçada.

As Figs. 7.2 e 7.4 sugerem um aparente escurecimento na imagem realçada, especificamente, em regiões onde os pixels possuem baixa intensidade. A imagem de entrada aparenta uma distribuição da luz mais uniforme. No entanto, esta percepção decorre da maior presença do *color cast* e turbidez, que caracterizam o efeito de *backscattering*, onde a luz ambiente refletida pela partículas em suspensão na água concorrem com a informação provinda dos objetos. Com o resgate da cor e, principalmente, contraste, há a remoção ou redução do *color cast*. Esta remoção causa percepção de diminuição da luz ambiente. No entanto, o que ocorre é a redução da componente de *backscattering* que atinge a câmera, que contém apenas informação de turbidez. Este efeito aparenta ser mais intenso em regiões com pixels de baixa intensidade, que tendem a representar níveis de refletância menores e próximos ao contexto real.

Complementando esta seção, são apresentados três grupos adicionais de imagens mostrando contextos diferentes relacionados à turbidez, aplicação e ação de *dehaze*. A Figura 7.5 mostra

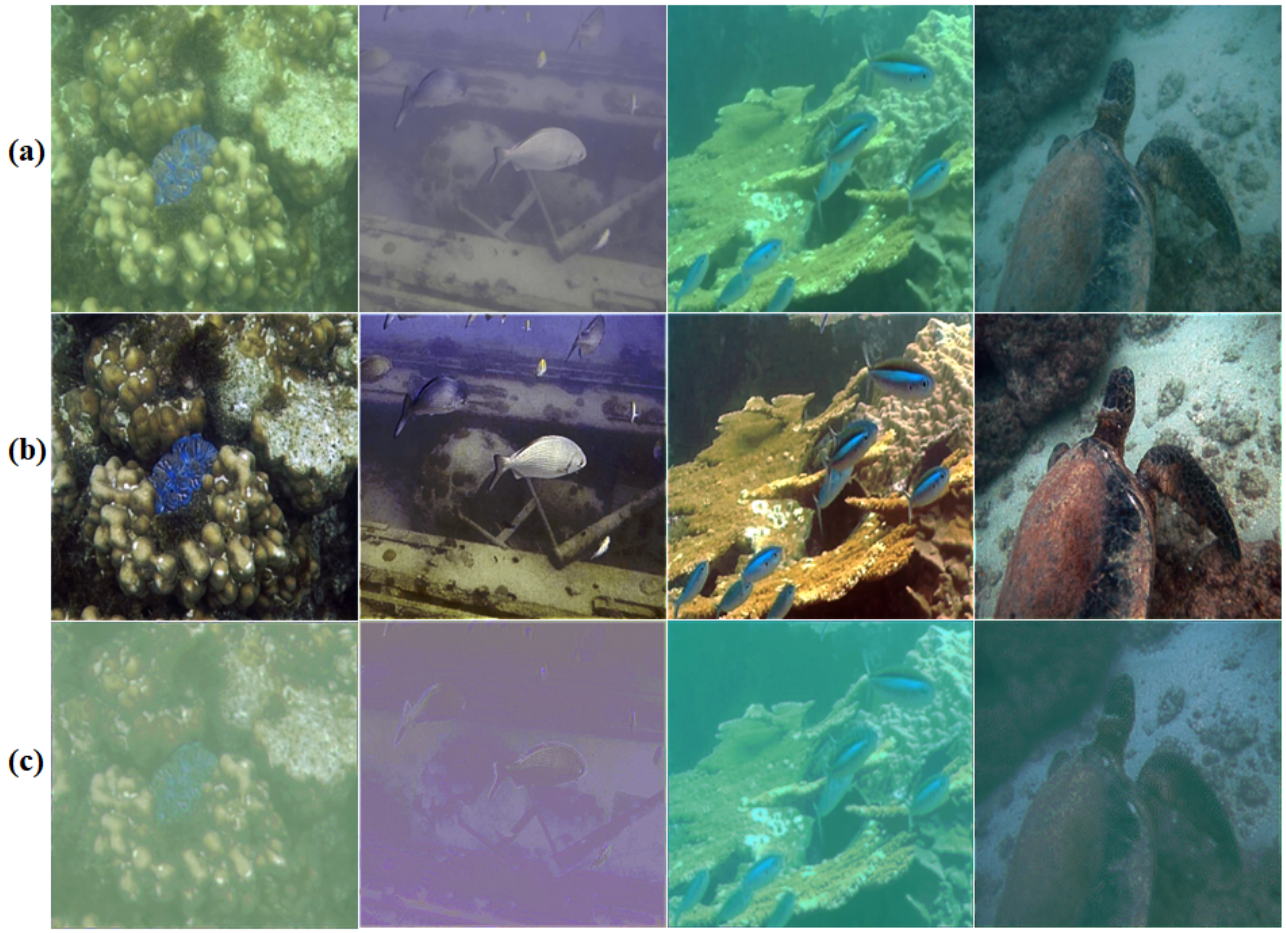


Figura 7.4: Exemplos de imagens de entrada com objetos próximos à câmera e resultantes da função de degradação e saída da rede. (a) Imagem de entrada $I_{(k)}$, (b) Imagem de saída da rede $\hat{I}_{(k)}$, (c) Imagem de saída do módulo de degradação $\hat{I}^{*}_{(k)}$. Imagens obtidas ao final do treinamento da rede.

conjunto de imagens com as cenas em ambiente de baixa turbidez. O modelo realiza realce, destacando cores e, principalmente, texturas e contornos dos objetos. A Fig. 7.6 mostra contexto voltado à aplicação. São mostradas imagens coletadas por ROV das instalações da usina nuclear Fukushima Daiichi, destruída por terremoto seguido de *tsunami* em 2011. Estas imagens não integraram o *dataset* usado no treinamento da rede. Apesar da baixa resolução e artefatos presentes nas imagens originais, o modelo mostra bom desempenho e realiza o realce de detalhes de cor e textura. Estas imagens estão disponibilizadas livremente na internet.

A Fig. 7.7 apresenta o resultado do modelo executando *dehaze*, que caracteriza a redução da presença de névoa, fumaça ou neblina em imagens adquiridas em ambiente atmosférico. São mostradas cenas sujeitas à presença de névoa ou neblina. O efeito destes fatores sobre as cenas é semelhante ao da água em imagens submersas, particularmente no ofuscamento dos objetos e perda de contraste. Entretanto, diferentemente de imagens UW, imagens em ambiente atmosférico tendem a apresentar faixa dinâmica unitária. Esta condição reduz o incremento de degradação gerada pela módulo de degradação, principalmente nas etapas envolvendo alterações no histograma e reforço/atenuação de cor e contraste, performadas na geração das imagens-

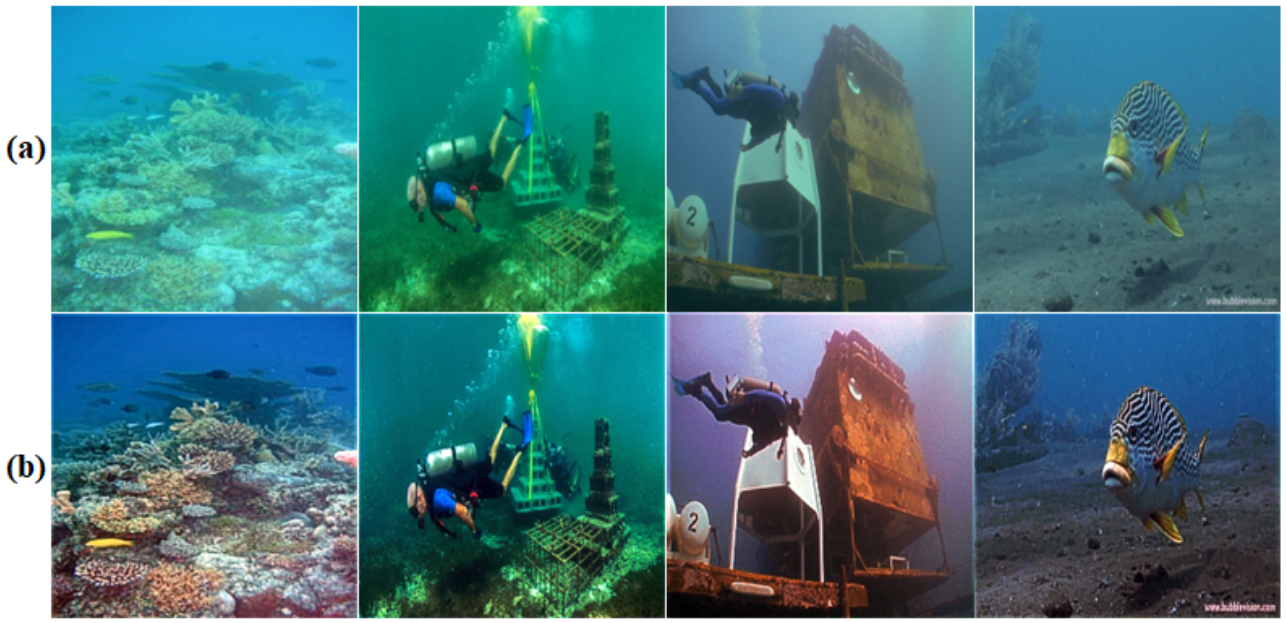


Figura 7.5: Exemplos de imagens de entrada com baixa turbidez. (a) Imagem de entrada, (b) Imagem de saída da rede.

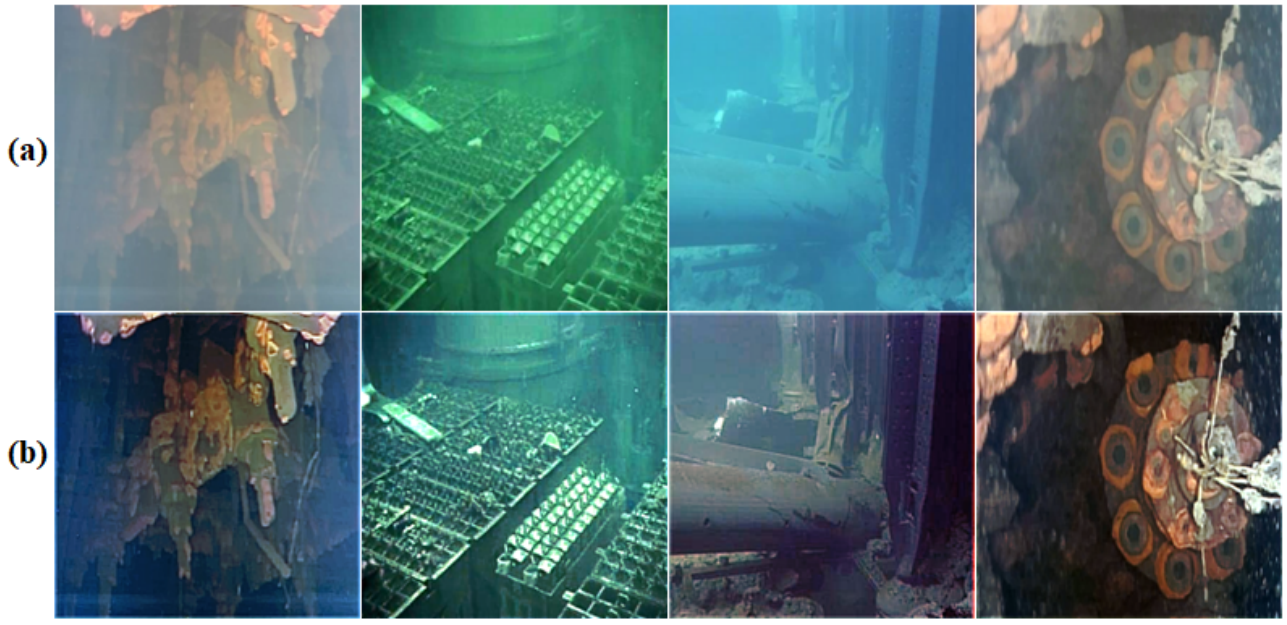


Figura 7.6: Imagens das instalações da usina nuclear Fukushima Daiichi - Japão. (a) Imagem de entrada, (b) Imagem de saída da rede.

guias para realce e degradação. No entanto, apesar de não ser desenvolvido para *dehaze*, o modelo mostra desempenho satisfatório e consegue reduzir o efeito da neblina, realçando bordas, contornos e textura dos objetos. Outro ponto a destacar, os resultados da Fig. 7.7 foram obtidos com o modelo treinado exclusivamente com imagens UW. Nenhum contexto envolvendo a presença de *haze* (neblina, poeira, fumaça) foi apresentando previamente à rede, assim como, nenhuma estratégia para ajuste do modelo (*fine tunning*) foi executada.



Figura 7.7: Imagens em ambiente atmosférico com cenas sujeitas à ação de neblina. (a) Imagem de entrada, (b) Imagem de saída da rede.

7.1.2 Estudo comparativo

Ao longo dos últimos anos, a melhoria de imagens subaquáticas têm recebido grande atenção dos pesquisadores, gerando um número expressivo de trabalhos publicados. Neste contexto, é importante estabelecer uma perspectiva de desempenho do modelo proposto a partir do estudo comparativo com métodos consolidados e pertencentes ao estado-da-arte.

Na avaliação comparativa, foram considerados modelos baseados em DL propostos por Li et al. (2021), Li et al. (2020), Islam et al. (2020), Saleh et al. (2022) e versão anterior do método proposto, apresentada em Mello et al. (2022). Em conjunto, também considerou-se os trabalhos de Ancuti et al. (2012) e Drews-Jr et al. (2016) baseados em processamento digital da imagem.

Na avaliação, os métodos são identificados por "Li2021"(Li et al., 2021), "Li2020"(Li et al., 2020), "Islam"(Islam et al., 2020), "Fusion"(Ancuti et al., 2012), "UDCP"(Drews-Jr et al., 2016), "Saleh"(Saleh et al., 2022), "Mello2022"(Mello et al., 2022) e o presente modelo como "Proposto".

Avaliação qualitativa

Neste contexto de avaliação, a análise é subjetiva e baseada na percepção visual, onde são considerados como critérios de qualidade a habilidade do método em recuperar cor e contraste, e a efetividade em reduzir *color cast* e o efeito da turbidez. Em conjunto, é considerada a constância de cores nas imagens resultantes, através da preservação e coerência das cores entre a imagem de entrada e a realçada.

A Fig. 7.8 apresenta amostras de imagens resultantes dos métodos considerados no estudo. Selecionou-se imagens que apresentam turbidez moderada, com diversificação na coloração e intensidade do *color cast*.

Embora os métodos mostrem ação de realce, os desempenhos são bastante irregulares. A maioria dos métodos reduz a percepção da turbidez. No entanto, a redução é muito limitada em Li2020 e Islam, onde percebe-se distorção de cor com forte realce associado ao canal vermelho. O método UDCP mostra efetividade na redução da turbidez, mas com severa distorção de cores e escurecimento da cena em algumas imagens. O método Fusion é bastante efetivo na redução da turbidez e recuperação de contraste. Entretanto, aparenta tendência à intensificar a "texturização"e produz artefatos em regiões de baixa frequência nas imagens. Em adição, sugere excessiva recuperação de cor em imagens com forte desequilíbrio entre os canais de cores.

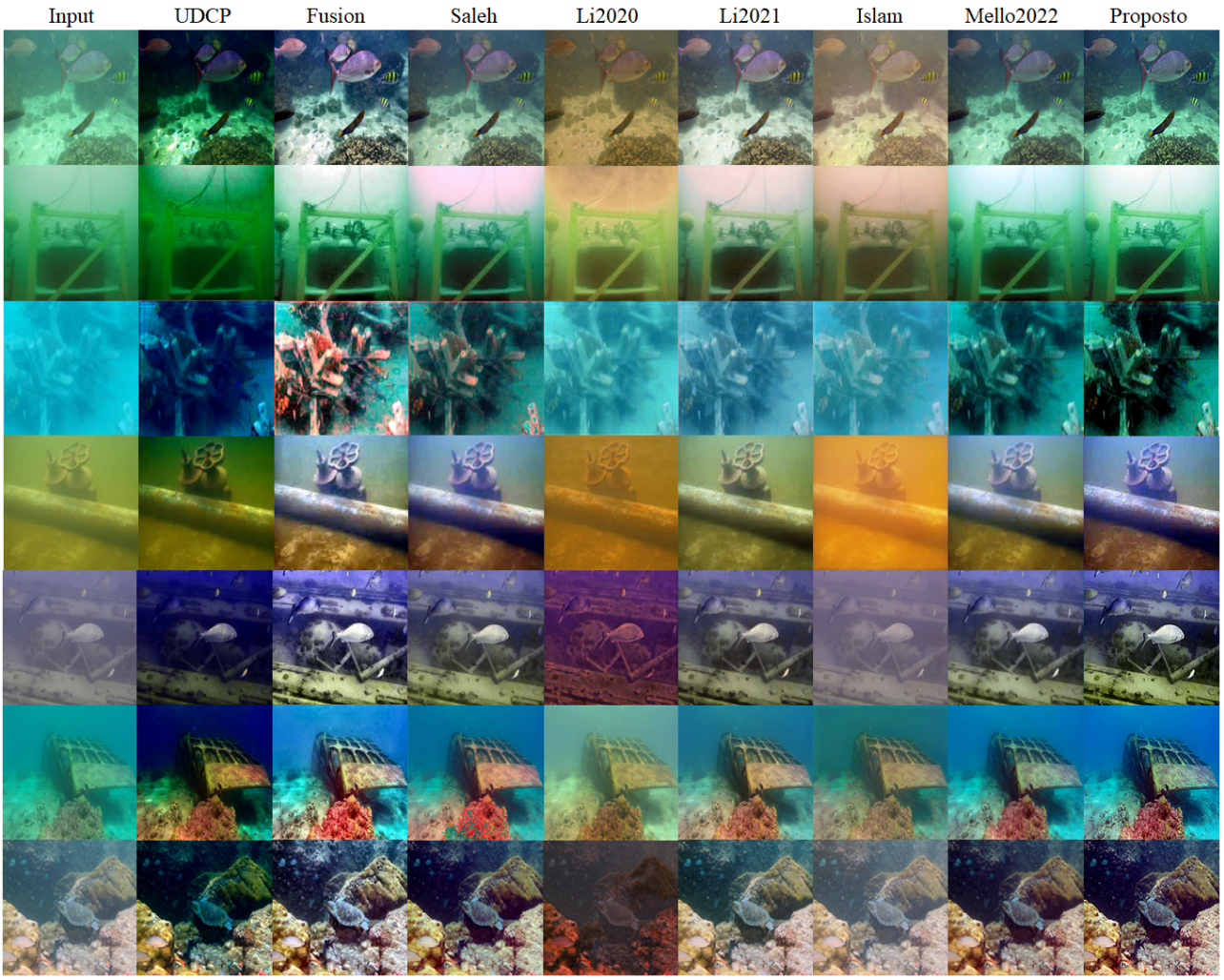


Figura 7.8: Exemplos de imagens resultantes dos métodos utilizados no estudo comparativo.

Os métodos Li2021, Saleh, Mello2022 e o proposto apresentam desempenho similares no realce de cor, contraste e redução da turbidez. Li2021 aparenta uma naturalidade maior nas cores, obtida a partir do treinamento utilizando dados pareados. Neste, as imagens de referência oferecem *features* adicionais ao modelo, permitindo que cores diversas às presentes na imagem de entrada sejam inferidas como resultado do aprendizado da rede. O modelo proposto mostra melhor desempenho no realce de contraste e na definição de bordas e contornos. Comparando com a expressiva recuperação de textura de Fusion, o método confere maior suavidade e equilíbrio na apresentação de cores e redução do borramento.

No caso específico da comparação entre a versão anterior Mello2022 com a atual, proposta, pode-se perceber nas imagens resultantes a melhoria na redução da turbidez e no realce, exercida pela ação dos módulos de textura e suavização que a versão anterior não conta.

Subjacente à análise, destaca-se a importância do *dataset* UWData utilizado no trabalho, que consiste em fator adicional para o bom desempenho do método proposto. A opção por mesclar imagens provenientes de *datasets* elaborados para diferentes objetivos (realce, segmentação, avaliação, identificação, classificação) na sua elaboração mostrou-se bastante efetiva. Na seleção das imagens, procurou-se valorizar os aspectos (*features*) da degradação sob diferentes contextos e intensidades com intuito de fortalecer o aprendizado da rede na redução ou remoção destas *features*.

Avaliação quantitativa

A avaliação quantitativa de imagens subaquáticas realçadas é uma questão que requer investigação, dado a carência de métricas específicas e a falta de consenso entre às existentes, quanto à concepção sobre qualidade de imagem. Outro aspecto que impacta nesta avaliação é a dificuldade com a virtual inexistência de conjuntos de dados com imagens de referência UW não sintetizadas. Estes fatores têm conduzido os autores a optarem por empregar uma associação de métricas, dispersas entre as específicas para ambiente subaquático e alternativas emprestadas da avaliação de imagens não-subaquáticas, requerendo ou não imagens de referência. Quando há a necessidade de imagens de referência, os autores recorrem à sintetização de imagens realçadas, cuja qualidade está associada ao método utilizado para obtê-las e que irá balizar o resultado (*score*) proposto pela métrica.

O método proposto não requer imagens de referência. Assim, foi definida estratégia para oferecer um contexto quantitativo para a avaliação que minimize a necessidade de se dispor de imagens de referência. Portanto, a avaliação é dividida em dois momentos, o primeiro contendo métricas sem referência (métricas cegas), onde são utilizadas métricas propostas para análise de imagens subaquáticas, especificamente, as métricas *Underwater Image Color Evaluation* (UCIQE) proposta por (Yang and Sowmya, 2015), *Underwater Image Quality Metric* (UIQM) proposta por (Panetta and Gao, 2015), e CCF apresentada em Wang et al. (2018). UCIQE avalia a degradação da cor no espaço de cores CIELab, baseada no desvio padrão da crominância, saturação média e diferença entre valores extremos da luminância. UIQM avalia a degradação da cor a partir da medida da coloração usando a teoria da cores opostas e o borramento avaliado com medidas de nitidez e contraste. Em ambas estas métricas, elevados *scores* indicam boa qualidade da imagem. A métrica CCF quantifica a degradação da cor com índice de colorização, e o borramento e falta de visibilidade com medida de contraste e índice de turbidez. Alto valor para o *score* CCF indica boa qualidade da imagem.

A Tabela 6 apresenta os resultados para as métricas sem referência UIQM, UCIQE e CCF relativas aos métodos comparados.

A observação dos resultados ilustra as diferentes perspectivas quanto à qualidade da ima-

gem oferecida pelas métricas, remetendo à falta de consenso comentada acima. Métodos com resultados menos expressivos para UIQM, apresentam as melhores desempenhos para UCIQE. Neste panorama, uma análise apenas à luz dos *scores* é prejudicada. Os valores resultantes são relacionados às *features* valorizadas no algoritmo de cada métrica, como saturação, contraste e crominância. Via de regra, as métricas não são sensíveis ao realce, ou seja, a avaliação entre a informação original e sua versão realçada não é valorizada e o quanto se agrega em qualidade na ação de realce não é considerado.

No entanto, como alternativa para a mitigação destas limitações, pode-se estabelecer uma comparação a partir da regularidade do desempenho dos métodos em cada métrica. Ou seja, considera-se que a classificação (*ranking*) dos desempenhos dos métodos em cada métrica sejam similares. Espera-se, assim, que um método com bom desempenho não mostre efeito "gangorra" em seus *scores*, com resultado bom ou muito bom em uma métrica e em outra seja ruim ou muito ruim. Assim, da Tab. 6 observa-se que o método proposto, junto com Mello2022 e Li2021 mostram melhor regularidade no desempenho ao longo das métricas. Pode-se destacar também o desempenho do método Fusion, embora apresenta forte oscilação entre os resultados para UIQM e o conjunto UCIQE/CCF. Esta linha de análise e comparação é coerente com a interpretação obtida da avaliação qualitativa, baseada na percepção visual dos resultados.

Importante destacar o emprego de imagens do *dataset* UWData na avaliação do desempenho dos métodos. Com exceção dos métodos UDCP e Fusion que são agnósticos com relação à base de dados UW utilizada, os demais métodos são baseados em abordagens por aprendizado profundo. Os modelos utilizados na obtenção dos resultados foram os treinados e disponibilizados na internet pelos respectivos autores. Entretanto, verifica-se que os*datasets* utilizados nos treinamentos destes modelos são os mesmos utilizados para a seleção de imagens na composição do UWData. Embora, haja coincidência parcial entre os *datasets* usados no treinamento de um modelo em comparação a outro, todos foram treinados com um conjunto de imagens comuns ao UWData. Esta condição subsidia favoravelmente às avaliações qualitativa e quantitativa, pois reduz a possibilidade de viés no desempenho devido à diferenças nos dados de treinamento.

Tabela 6: Resultados para as métricas UIQM, UCIQE e CCF. UDCP (Drews-Jr et al., 2016), Fusion (Ancuti et al., 2012), Mello2022 (Mello et al., 2022), Li2020 (Li et al., 2020), Li2021 (Li et al., 2021) e Islam (Islam et al., 2020).

Metric	UIQM ↑	UCIQE ↑	CCF ↑
UDCP	1,604	**6,915**	34,229
Fusion	2,623	5,568	**36,895**
Saleh	2,255	4,369	22,694
Mello2022	2,649	4,623	23,492
Li2020	2,654	3,281	13,737
Li2021	**2,913**	3,593	19,824
Islam	2,628	3,414	14,282
Proposto	2,726	4,745	30,628

A avaliação quantitativa é complementada com o emprego de métricas que requerem imagens

de referência, utilizadas regularmente na avaliação de imagens. Procura-se, também, verificar a validade das conclusões obtidas da análise com métrica sem referência.

Nesta etapa, são utilizadas as métricas *Peak Signal-to-Noise Ratio* (PSNR), *Mean Square Error* (MSE), *(*Structural Similarity) (SSIM) (Wang et al., 2004), *Gradient Magnitude Similarity Deviation* (GMSD) (Xue et al., 2013), CIEDE2000 (Sharma et al., 2005), *Feature Similarity* e *Feature Similarity with Chrominance* (FSIM and FSIMc) Zhang et al. (2011). Especificamente, PSNR e MSE são métricas associadas à qualidade da reconstrução da imagens, SSIM e GMSD são associadas ao contexto estrutural, CIEDE2000 vinculada à cor e FSIM e FSIMc ao aspecto da percepção visual.

A avaliação é realizada utilizando as imagens do conjunto Milk do *dataset* Turbid, Duarte et al. (2016). O conjunto contém 20 imagens da mesma cena com níveis crescentes de turbidez, sendo que a primeira imagem foi adquirida com a água sem turbidez (#100). A imagem de referência utilizada na aplicação das métricas é derivada desta imagem com água sem turbidez. Ela consiste de uma versão com histograma expandido e balanço de cor (*color balance*). A opção por estes ajustes ocorreu em função desta imagem apresentar faixa dinâmica menor que a unidade e ter sido adquirida sob iluminação controlada como as demais imagens do Milk, o que conferiu aspecto escurecido às cenas. Este escurecimento é interpretado como degradação pelos métodos no estudo comparativo, gerando versões realçadas, de melhor percepção visual que a imagem original sem turbidez do Milk. Este resultado inviabiliza o uso direto da imagem 100 como referência, pois acarretaria em incerteza quanto *scores* das métricas, dado que as imagens realçadas apresentariam aspecto qualitativo melhor que a referência. As Figs. 7.9(a) e (c) apresentam, respectivamente, a imagem Milk #100 e a versão utilizada como imagem de referência no estudo avaliativo com métricas referenciadas. Em conjunto, são mostrados as respectivas distribuições no espaço de cores de ambas as imagens.

A versão com histograma expandido e equilíbrio de cor, propicia melhor percepção de cor e contraste e reduz o aspecto escurecido que a imagem original mostra. A distribuição dos pixels mostra diversificação das cores ampliada e a imagem descrevendo volume maior no espaço RGB.

Na Fig. 7.10 mostra as imagens resultantes dos métodos para amostras de imagens do *dataset* Turbid/Milk. A Tab. 7 complementa a avaliação com os *scores* para as métricas com referência descritas acima.

A partir da perspectiva visual, observa-se relativa efetividade dos métodos para as imagens com baixa turbidez (até #108). Para as imagens com nível acentuado de turbidez, os métodos Islam, Li2020 e Li2021 expõe baixo desempenho, com produção de artefatos e dificuldades na recuperação cores, de boras e contornos. Os métodos Fusion, Mello2022 e o proposto mostram resultados muito similares quanto ao tratamento do *color cast*. No entanto, a medida que a nível de degradação aumenta, Mello2022 aparenta perda de efetividade na redução da turbidez e Fusion sofre com a incremento do ruído de fundo e geração de artefatos associados, mostrando também redução de cor. O método proposto mostra melhor habilidade no tratamento da turbidez e manutenção da cor e contornos dos objetos.

Os resultados na Tab. 7 sugerem concordância com a análise subjetiva, com os métodos

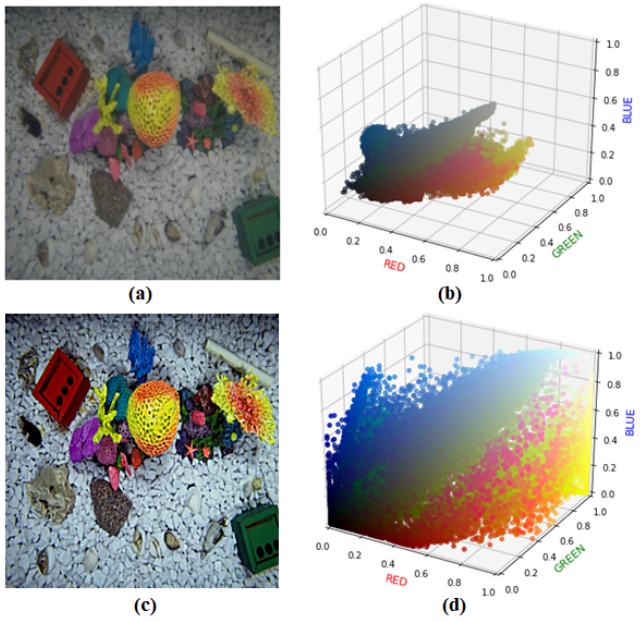


Figura 7.9: Imagem do *Turbid/Milk* com água sem turbidez significativa e a versão utilizada como imagem de referência no estudo comparativo. (a) Imagem Milk#100, (b) Distribuição dos pixels no espaço de cores de Milk#100, (c) Imagem utilizada como referência, e (d) Distribuição dos pixels no espaço de cores da imagem de referência.

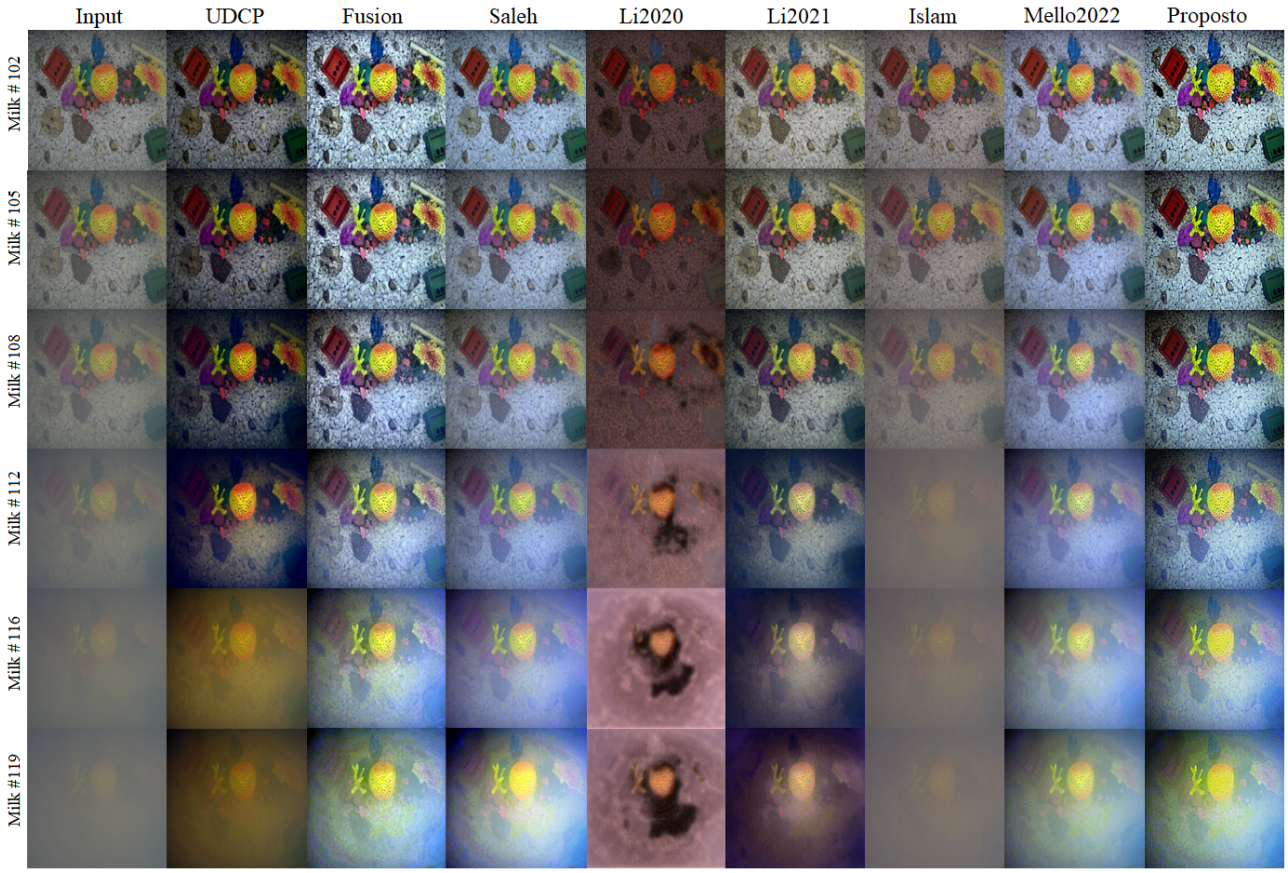


Figura 7.10: Imagens resultantes dos métodos para imagens do *dataset* Turbid/Milk.

Tabela 7: Resultado quantitativo para as métricas com referência

Métrica	PSNR ↑	MSE ↓	SSIM ↑	CIE2K ↓	FSIM ↑	FSIMc ↑	GMSD ↓
UDCP	10.64	0,089	0,480	27,97	0,970	0,696	$50e^{-6}$
Fusion	16,70	**0,027**	0,673	15,00	0,975	0,800	$32e^{-6}$
Saleh	9,785	0,107	0,489	35,59	0,973	0,632	$54e^{-6}$
Mello2022	16,52	0,039	0.632	17,75	**0.979**	0.714	$39e^{-6}$
Li2020	11,29	0,083	0,428	30,56	0,924	0,600	$69e^{-6}$
Li2021	14,04	0,051	0,595	21,52	0,973	0,735	$42e^{-6}$
Islam	14,39	0,038	0,494	19,41	0,974	0,581	$59e^{-6}$
Proposto	**17,25**	**0,027**	**0,723**	**14,57**	0,973	**0,819**	$\mathbf{28}e^{-6}$

Fusion, Mello2022 mostrando desempenho similar nas métricas. Interessante destacar que as maiores diferenças entre os métodos Fusion e o Proposto ocorrem nas métricas associadas ao ruído (PSNR) e à cor (CIE2K e FSIMc), com favorecimento para o método proposto. Isto também corrobora as observações suscitadas da avaliação subjetiva quanto à perda na cor e à intensificação de ruído no Fusion. O desempenho satisfatório nas métricas associadas à ruído e estrutura (PSNR, MSE, SSIM), em conjunto com o obtido relativo à percepção de *features* em escala de cinza (FSIM), são importantes para a percepção de imagens subaquáticas. Pois, como indicado por Burger and Burge (2016), a visão humana é menos sensível à mudanças na cor do que à luminância. No ambiente aquático, os efeitos da profundidade e a turbidez sobre as cores são acentuados sob luz ambiente reduzida, condição bastante comum em atividades submersas.

O mapa de gradientes em uma imagem podem indicar o nível de definição de bordas e contornos e são sensíveis à degradação. Diferentes estruturas locais em uma imagem distorcida sofrem diferentes graus de degradação. A métrica GMSD captura a qualidade local, determinando o desvio padrão do mapa de gradientes da imagem como índice de qualidade da imagem. O método proposto mostra bom desempenho para GMSD, indicando que as imagens resultantes permitem boa percepção do conteúdo de alta frequência na imagem, como bordas e contornos.

No âmbito dos métodos baseados em DL, a metodologia proposta mostra bom desempenho, considerando que não utiliza dados pareados, módulos pré-treinados ou pré-processamento das imagens de entrada. Entretanto, não considera-se esta condição como uma limitação, mas uma característica do método, dado que os tons e intensidades reais das cores em uma cena subaquática são desconhecidas na grande maioria dos casos.

7.1.3 Resultados com o módulo de atenção para áreas saturadas

As limitações na melhoria de imagens provocadas na expansão do histograma foram tratadas a partir do uso do módulo de atenção proposto. O efeito do módulo de atenção sobre amostras de imagens com desequilíbrio entre os canais de cores é mostrado na Fig. 7.11. Na Fig. 7.11(c) são mostradas as imagens referentes à $I_{at(k)}$ para cada caso. Em função do canal vermelho ser o mais facilmente atenuado no ambiente subaquático, $I_{at(k)}$ apresenta informação

a ser compensada com predomínio da cor vermelha na maioria dos casos. No entanto, este desequilíbrio é o resultado das condições do ambiente em que a cena se encontra e outras colorações são podem ser dominantes em $I_{at(k)}$. Como exemplo, a imagem mostrando algas em ambiente de intensa turbidez e *color cast* de tom amarelo, resultou na respectiva $I_{at(k)}$ com predominância do canal azul.

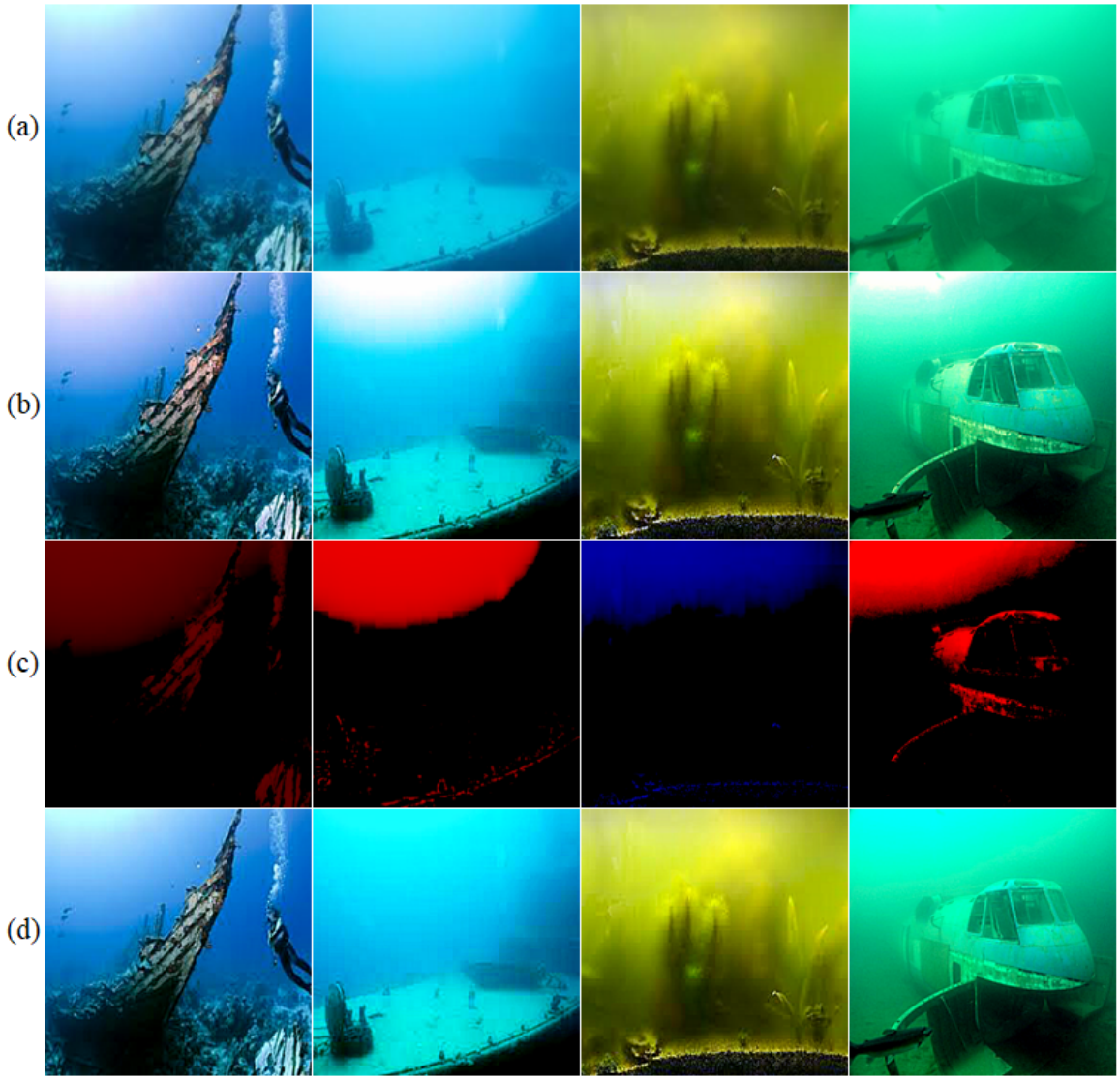


Figura 7.11: Amostras de imagem com desequilíbrio entre os canais de cores atenuados pelo módulo de atenção. (a) Imagem original $I_{(k)}$, (b) Imagem realçada $\hat{I}_{(k)}$ sem o módulo de atenção no *pipeline*, (c) Imagem resultante do módulo de atenção $I_{at(k)}$ e (d) Imagem realçada $\hat{I}_{(k)}$ com o módulo de atenção no *pipeline*.

A análise visual das imagens na Fig. 7.11(d) sugere redução na intensidade da ação de realce, especificamente, nas áreas saturadas ou com alta intensidade dos pixels. Nas demais regiões da imagem, o realce aparenta mesma intensidade. Entretanto, esta alteração no resultado para o realce impacta na qualidade da imagem. Na Tab. 8 são comparados os resultados para as métrica sem referência UIQM, UCIQE e CCF para o modelo com e sem o módulo de atenção. As métricas UIQM e CCF mostram redução no *score*, enquanto UCIQE registra

aumento. Novamente, as distintas interpretações sobre qualidade da imagem UW expressas pelas métricas conduz a diferentes resultados.

Tabela 8: Resultado quantitativo para as métricas sem referência UIQM, UCIQE e CCF. Modelos sem e com o módulo de atenção são indicados por (Mod) e (Mod+Att), respectivamente.

	Mod	Mod+Att
UIQM ↑	2,726	2,535
UCIQE ↑	4,745	5,181
CCF ↑	30,628	28,729

Embora as imagens do Turbid/Milk não apresentem desequilíbrio entre os canais de cores, realizou-se a avaliação envolvendo as métricas que requerem imagem de referência. O objetivo é estender a avaliação sobre o impacto do módulo de atenção sobre a ação de realce do modelo. Os resultados são mostrados na Tab. 9. Semelhante aos resultados para métricas cegas, o efeito sobre o realce do ponto de vista da qualidade de imagem apresenta baixo impacto. Isto decorre do fato de que a atuação do módulo de atenção é mais intensiva nas regiões mais afetadas pelo desequilíbrio dos canais de cores, mantendo o restante da imagem sem alterações.

Tabela 9: Resultado quantitativo para as métricas com imagem de referência PSNR, MSE, SSIM, CIEDE2000 (CIE2K), FSIM, FSIMc, and GSMD. Modelos sem e com o módulo de atenção são indicados por (Mod) e (Mod+Att), respectivamente.

	Mod	Mod+Att
PSNR ↑	17,25	16,92
MSE ↓	0,027	0,032
SSIM ↑	0,723	0,709
CIE2K ↓	14,57	15,97
FSIM ↑	0,973	0,970
FSIMc ↑	0,819	0,817
GSMD ↓	$28e^{-6}$	$28e^{-6}$

7.2 Resultados para o realce de vídeo

O realce de vídeos subaquáticos requer além do resgate de cor e contraste que a ação de realce apresente constância nos resultados de um *frame* para outro, ou seja, mostre consistência temporal.

No algoritmo para realce proposto, são utilizadas as operações de expansão e contração do histograma. Como discutido anteriormente na seção 5.2, estes procedimentos são sensíveis à mudanças nos valores máximos e mínimos do *frame*, que podem variar consideravelmente de um quadro para outro. Isto resulta em diferentes ações de realce. O método proposto processa estas variações a partir de uma abordagem centrada na média, de maneira a prover consistência temporal à sequência de vídeo realçada. Nesta seção são apresentados e discutidos os resultados

da aplicação da metodologia para vídeo. Neste contexto, os aspectos associados ao realce e à consistência temporal estão implicitamente postos no vídeo realçado resultante.

A Fig. 7.12 mostra conjuntos de *frames* de vídeos e as respectivas versões realçadas utilizando-se a metodologia apresentada. Os vídeos originais estão disponíveis na internet, sendo identificados por "Octopus (Cams, 2018), "Isla Socorro" (Rojas, 2023), "Saint Martin" (Ernst, 2022), "Fish Farm" (Trecker, 2021). Estas sequências foram adquiridos a partir de diferentes fontes (webcam, ROV, mergulhadores), mostrando diferentes condições ambientais, luz, *color cast* e turbidez. Importante destacar que os vídeos selecionados são contínuos e sem alterações, ou seja, não apresentam cortes ou transições entre quadros com contextos diferentes, mudanças de cor realizadas artificialmente, legendas ou outros processos similares. O método não requer quadros ou imagens de referência e a informação utilizada é fornecida apenas pelo *frame* de entrada. Assim, artefatos resultantes de edição prévia da sequência de vídeo poderiam gerar erros de interpretação pelo algoritmo, conduzindo a resultados indesejáveis.

A observação visual das amostras da Fig. 7.12 mostra que o desempenho em termos de realce é similar ao obtido para imagens. O modelo proporciona redução no *color cast* e no efeito da turbidez. Nas cenas com realce há melhor definição de bordas e contornos, com manutenção da constância de cores e incremento na percepção do contraste. O modelo foi treinado com o GFR realizando o ajuste dos valores extremos superior e inferior dos *frames*, considerando a média de 1% dos pixels de maior e menor intensidades por canal de cor, respectivamente.

Na efetivação do realce, caraterísticas de cor e contraste são incrementadas no resultado. No contexto do método, artefatos de *flickering* são gerados por variações nos valores máximos e mínimos entre *frames* sucessivos, como discutido na seção 5.3. No entanto, o vídeo original pode apresentar oscilações naturais entre cenas sucessivas, definidas pelo movimento da câmera e objetos, mas, principalmente, da fonte de luz e a interação resultante com o movimento da água. Estas últimas podem ser extremamente sutis, mas quando realçadas simulam a presença de *flickering*, mas são *features* do vídeo original, efetivamente produzidas pela água. Neste caso estão incluídos os vídeos "Sonar" e "Saint Martin", adquiridos sob boa percepção de luz ambiente e turbidez moderada baixa. A observação detalhada do contexto das cenas nas sequências originais permite identificar estas variações. Nas versões realçadas, estes efeitos são parcialmente minimizados, dado que afetam a cena como um todo, impactando nos valores médios dos canais de cores e não apenas valores específicos como mínimos e máximos. Sendo, assim, de maior dificuldade para minimização.

A estratégia adotada para o treinamento é importante para o bom desempenho do modelo, sendo este realizado em duas etapas. Na primeira etapa, a rede neural é treinada com *dataset* composto apenas por imagens. O objetivo é permitir o mapeamento de *features* associadas ao ambiente subaquático e fortalecer a habilidade de generalização do modelo. Nenhuma sequência de vídeo é utilizada nesta etapa. Na segunda etapa, é realizado o *fine tunning* ao vídeo-alvo, onde é utilizado um conjunto contendo 20% de *frames* não-contínuos do vídeo. Esta etapa é importante na qualidade final do resultado, pois permite o ajuste do modelo ao contexto ambiental do vídeo. Em vídeos previamente adquiridos, esta etapa requer apenas a seleção do

conjunto de *frames* para o ajuste. Em todos os testes, esta segunda etapa do treinamento foi realizada a partir de conjuntos de *frames* não-contínuos de uma sequência de vídeo contínua contendo, no mínimo, 800 *frames*. Para uma potencial aplicação em tempo real, o *fine tunning* pode ser realizado a partir da coleta *in loco* deste conjunto de dados. Esta segunda etapa do treinamento requer um intervalo de tempo muito pequeno para ser efetivada. Isto representa uma característica importante para a aplicação. Por exemplo, a aquisição de 1 *frame* a cada 5 observados, a uma taxa de 25 quadros/segundo, permite a geração de 300 *frames* não-contínuos em 1 minuto de aquisição, condição suficiente para a efetivação da segunda etapa do treinamento. Assim, tarefas subaquática como monitoração ou exploração e inspeção baseadas em ROV, podem contar com o modelo pré-treinado na primeira etapa e realizar este *fine tunning* imediatamente antes do início da missão.

Trechos dos vídeos citados e outros exemplos realçados com a metodologia proposta, estão disponíveis para visualização[3] na internet.

[3] `https://drive.google.com/drive/folders/1nEdm2SsHCwHs9Opu_KEYzZMrQvRGc3cS?usp=drive_link`

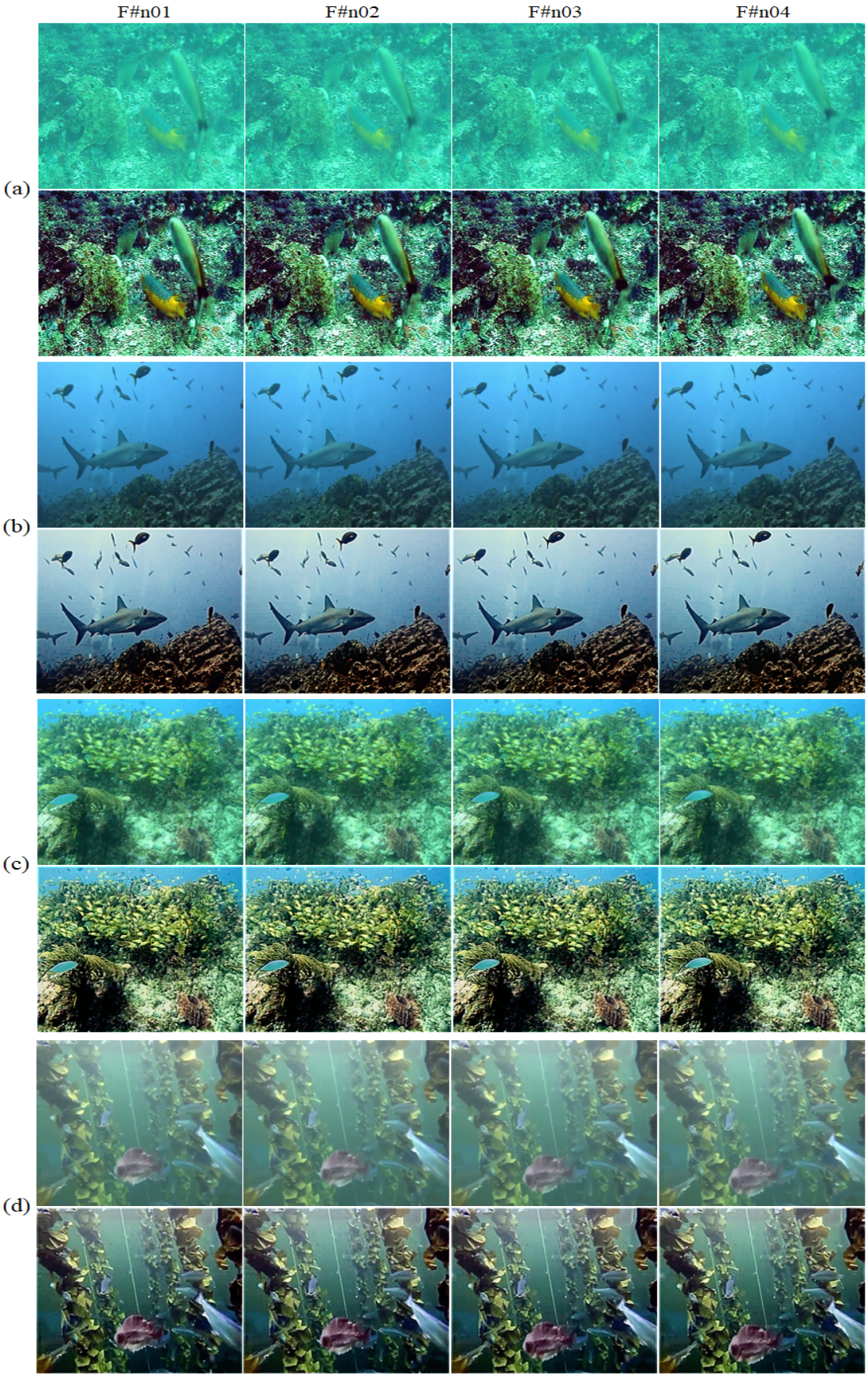


Figura 7.12: Frames de vídeos com respectivas versões realçadas. (a) Vídeo: "Octopus, (b) Vídeo: "Isla Socorro", (c) Vídeo: "Saint Martin" e (d) Vídeo: "Fish Farm".

7.2.1 Estudo comparativo envolvendo métodos para realce subaquático

Complementando a análise subjetiva da seção anterior, desenvolveu-se estudo comparativo entre o método proposto e os métodos propostos por Islam et al. (2020), Li et al. (2020) e Saleh et al. (2022) abordados na avaliação envolvendo imagens. Estes trabalhos foram apresentados originalmente para o realce de imagem, entretanto, seus autores realizaram avaliações dos respectivos métodos para o realce de vídeo. Em conjunto, utilizamos versão prévia do modelo apresentada em Mello et al. (2022), dado que apresentou bom desempenho junto aos métodos citados no contexto de imagem.

As Figuras 7.13 e 7.14 mortram os *frames* resultantes para os métodos citados e o proposto. Para esta comparação foram considerados os vídeos "Los Arcos" (Detloff, 2022) e "Sonar" (Ryan, 2021). Estes vídeos foram selecionados em função de apresentarem distância relativa dos objetos à câmera e níveis de turbidez (moderado) semelhantes, permitindo uma análise concisa dos desempenhos dos métodos. Estes vídeos apresentam ainda uma dinâmica associada ao movimento da câmera, incidência e posição relativa da fonte de luz bastante desafiadoras ao realce. Entretanto, não é possível a percepção destes detalhes nas figuras.

Os método são efetivos na ação de realce, onde o resgate de cor ocorre em todos os resultados. Entretanto, os métodos Islam e Li2020 sugerem fortalecimento na compensação do canal vermelho, com Islam mostrando bom desempenho em termos de suavidade e definição de bordas nos objetos. O desempenho quanto a redução da turbidez é limitado nestes métodos, impactando na percepção da textura. Os métodos Saleh e Mello2022 mostram desempenhos similares. Ambos os métodos utilizam a expansão do histograma como parte de seus algoritmos, conferindo tonalidades semelhantes às cenas. Saleh confere melhor contraste às cenas do vídeo "Los Arcos", resultando uma melhor percepção da textura, e sugere melhor uniformidade na ação de realce. No entanto, não mantém o mesmo desempenho para o vídeo "Sonar", onde Mello2022 alcança melhores resultados.

O método proposto mostra bom desempenho em ambos os vídeos. Promove a redução da turbidez e do *blurring*. Em consequência, há fácil percepção de textura, bordas e contornos, configurando o efeito das etapas relativas à suavização e textura nos *pipelines* de realce e degradação. O resgate de cor e contraste é mais evidente no vídeo "Sonar", com constância das cores mantida relação à informação de entrada, em ambos os vídeos.

A avaliação quantitativa para o realce de vídeos subaquáticos deve responder pelo resultado da ação de realce com consistência temporal. Esta questão torna-se bastante complexa, pois a percepção da melhoria deve ser a mesma ao longo do vídeo, assim estes dois aspectos devem ser considerados ao mesmo tempo sobre o vídeo realçado. Isto é fortalecido pelo fato da consistência temporal ser intrinsecamente dependente do resultado para realce. Entretanto, não há disponibilidade de métricas para vídeo que contemplem estes dois aspectos oferecidos para realce de vídeo UW. Em geral, as métricas para realce de vídeo não são orientadas para a aplicação subaquática, requerem sequências de referência ou utilizam datasets para comparação com sequências não-subaquáticas (Wu et al., 2022). Na busca por alternativa para

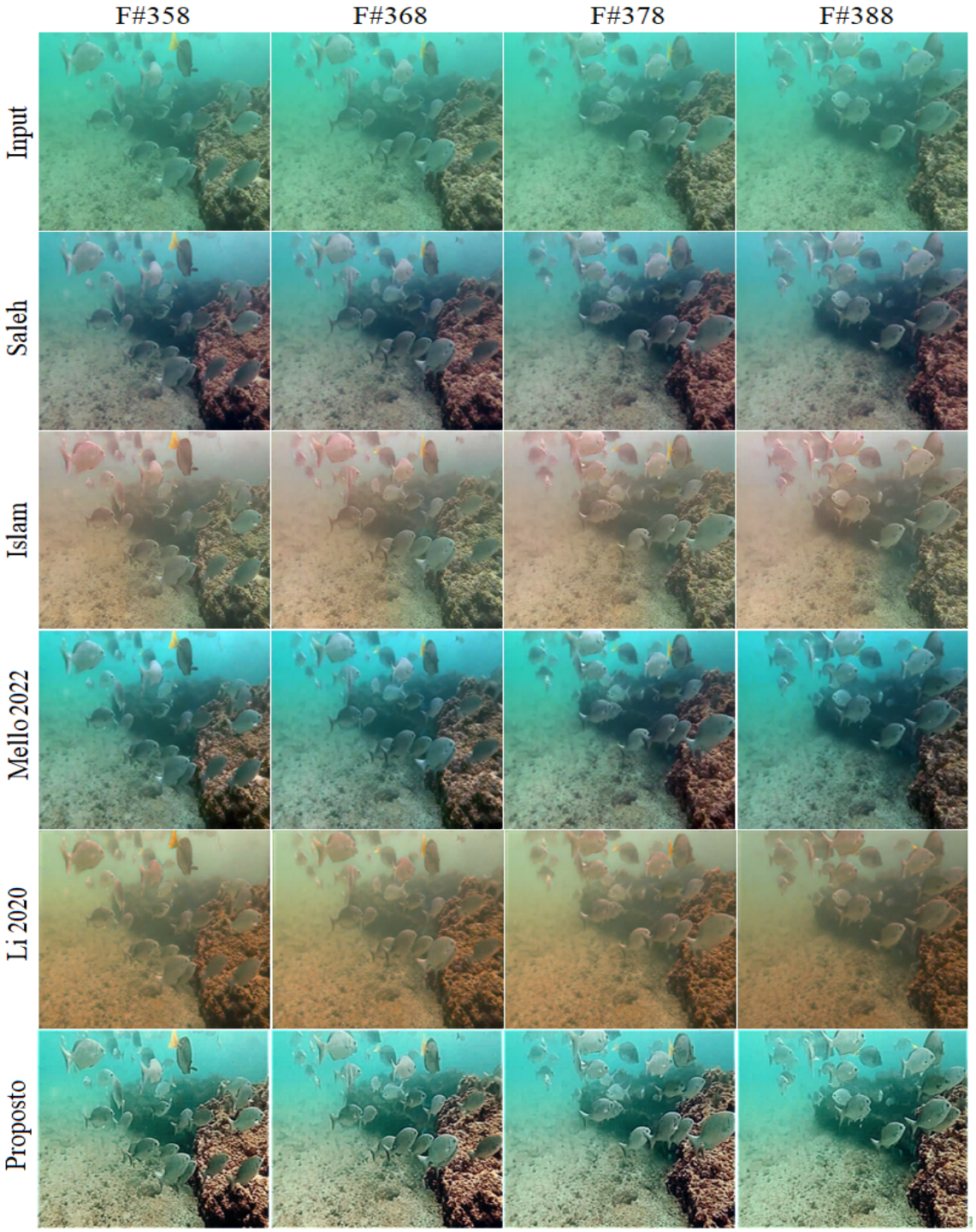


Figura 7.13: Resultado comparativo para o vídeo “Los Arcos” (Detloff, 2022). Métodos: Saleh (Saleh et al., 2022), Islam (Islam et al., 2020), Mello2022 (Mello et al., 2022), Li (Li et al., 2020) e o proposto.

permitir avaliação quantitativa, realce e consistência temporal são considerados em abordagens separadas.

Para a avaliação dos resultados na questão realce, a abordagem segue o proposto para imagens subaquáticas. Assim, foram utilizadas as métricas sem referência UIQM, UCIQE e CCF. Dado que não se dispõe de sequências de vídeo de referência, métrica referenciadas não

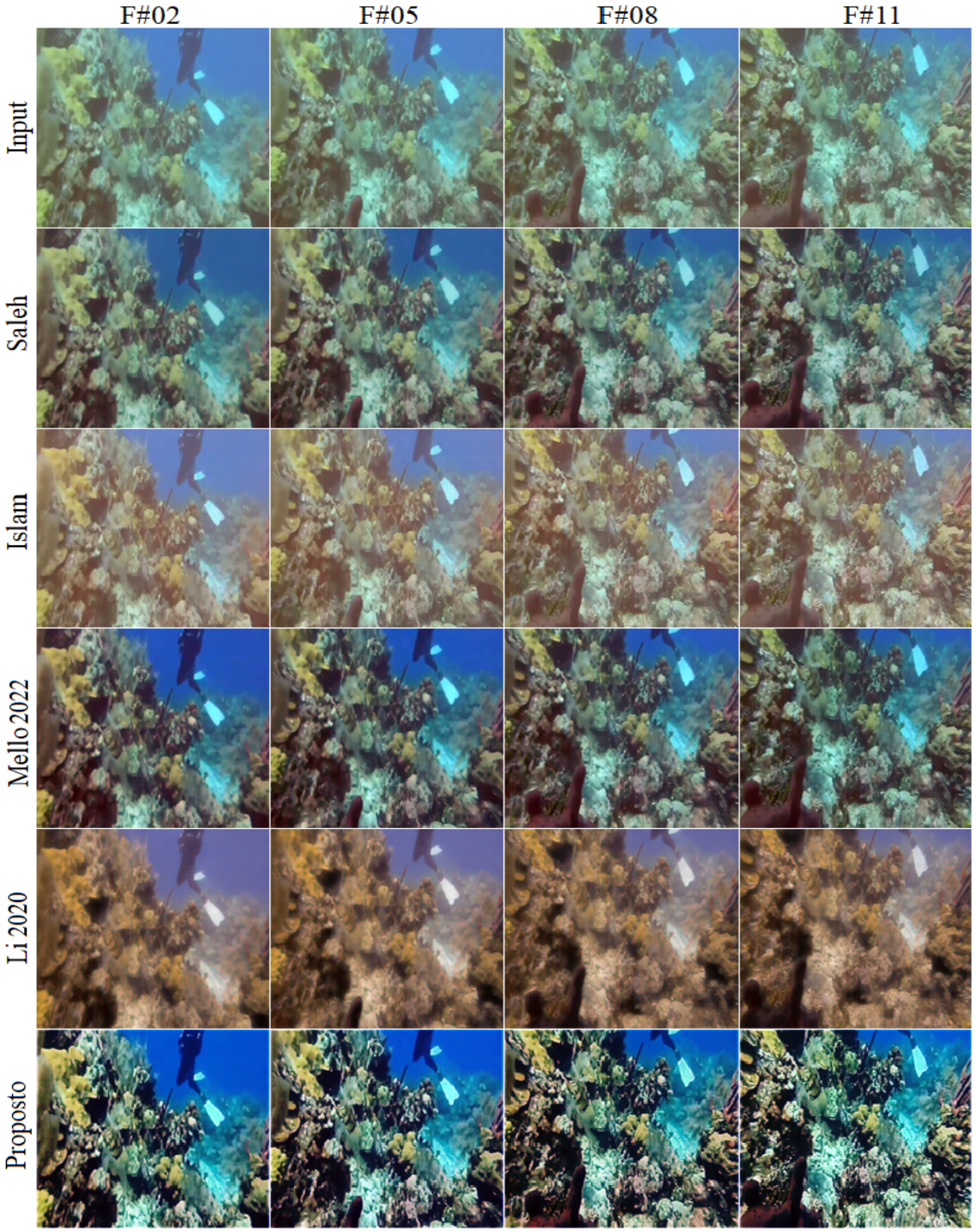


Figura 7.14: Resultado comparativo para o vídeo “Sonar” (Ryan, 2021). Métodos: Saleh (Saleh et al., 2022), Islam (Islam et al., 2020), Mello2022 (Mello et al., 2022), Li (Li et al., 2020) e o proposto.

foram consideradas.

As métricas destinadas à avaliação da consistência temporal para vídeos são, em sua maioria, orientadas à avaliação de algoritmos de compactação e requerem vídeos de referência para a comparação. Neste tipo de aplicação, o vídeo de entrada se constitui na própria referência, facilitando a avaliação. O método proposto não utiliza informação de referência e não se

dispõe de vídeos subaquáticos de referência para realce. Assim, optou-se para a avaliação da consistência temporal utilizar-se a métrica sem referência *Natural Image Quality Evaluator* (NIQE) apresentado por(Mittal et al., 2013). NIQE considera a sensibilidade da visão humana a áreas de alto contraste em uma imagem, usando distribuição normal multivariada para definir modelo para avaliação destas áreas sensíveis. Valores baixos para o *score* NIQE indicam boa qualidade da imagem. Entretanto, ao invés de realizar avaliação utilizando a abordagem quadro-a-quadro, como na avaliação de imagens, integra-se as estatísticas dos *frames* a nível de pixel e o *score* NIQE é definido para o vídeo como um todo, como proposto no trabalho de Saad et al. (2014). Esta abordagem permite a captura de aspectos associados à distorção temporal presente na sequência de vídeo. Na avaliação, esta metodologia é identificada por **NIQEv**. Em conjunto, emprega-se a variação temporal dos coeficientes médios (DC) obtidos da transformada discreta do cosseno (*discrete cosine transform* - DCT). As variações temporais na intensidade média em quadros sucessivos podem ser estimadas através da taxa de variação dos coeficientes no tempo. Para tanto, a média das variações é obtida através da expressão indicada na Eq. 7.1, ou seja:

$$DC_m = \frac{1}{N-1} \sum_{i=1}^{N-1} \frac{|D_{i+1} - D_i|}{D_{i+1} + D_i}, \tag{7.1}$$

onde DC_m é o valor médio da variação entre os coeficientes DC, N é o número de *frames*, D_{i+1} e D_i são os coeficientes DC dos *frames* indexados por $i+1$ e i, respectivamente. Assim, baixos valores para DC_{avg} indicam reduzido conteúdo de variações temporais. Esta abordagem para a avaliação é coerente com a premissa do trabalho que considera numericamente pouco significativa a variação do valor médio dos pixels de um quadro para outro. Ainda, as diferenças são normalizadas por $(D_{i+1} + D_i)$ como forma de avaliar-se apenas a questão temporal, obtendo um panorama agnóstico quanto a ação de realce desenvolvida por cada método. Ou seja, minimizar o efeito do realce sobre os valores absolutos das diferenças.

A Tab 10 relaciona os resultados para a avaliação quantitativa. Os valores indicados correspondem às médias entre os valores obtidos para os vídeos "Los Arcos" e "Sonar".

Tabela 10: Resultados quantitativos para vídeo.

Método	**UIQM↑**	**UCIQE↑**	**CCF↑**	**NIQEv↓**	**DC_m ↓**
Islam	**3,141**	3,949	34,683	10,840	0,277
Li	2,956	4,186	36,840	12,060	0,117
Mello	2,933	5,580	45,173	11,302	0.161
Saleh	3,032	4,280	43,339	11,071	0,323
Proposto	3,125	**6,351**	**59,568**	**9,609**	**0,097**

7.2.2 Resultados para vídeos adquiridos em ambiente subaquático com reduzida iluminação e de alta turbidez

Os vídeos adquiridos em atividades submersas tais como inspeção, busca & salvamento e monitoramento tem alta probabilidade de serem adquiridos em ambientes com baixa visibilidade

e alta turbidez. A potencial aplicação do método incentivou a avaliação do desempenho em ambientes com alta degradação. Com este objetivo, procedeu-se a aquisição de vídeos nas instalações da Estação Marinha de Aquacultura da Universidade Federal de Rio Grande (EMA) (-32.203969, -52.177845).

Na EMA é desenvolvida pesquisa associada a peixes da espécie "Miraguaia" (*Pogonias cromis*), relativo a criação em cativeiro e reprodução. Os peixes são mantidos em tanques circulares com dimensão aproximadas de 4m de diâmetro e 1,7m de altura. Nestes tanques é simulado o ambiente natural da espécie: água salgada, luz ambiente muito baixa e alta turbidez. Na Fig. 7.15 são apresentados conjuntos de *frames* de dois vídeos coletados no interior destes tanques, com as respectivas versões realçadas. Os vídeos são identificados por EMA_TQ1 e EMA_TQ2 podem ser visualizados na internet[4]. EMA_TQ1 foi obtido no interior de tanque abrigando exemplares adultos jovens e EMA_TQ2 em tanque com adultos reprodutores, de maior tamanho e peso.

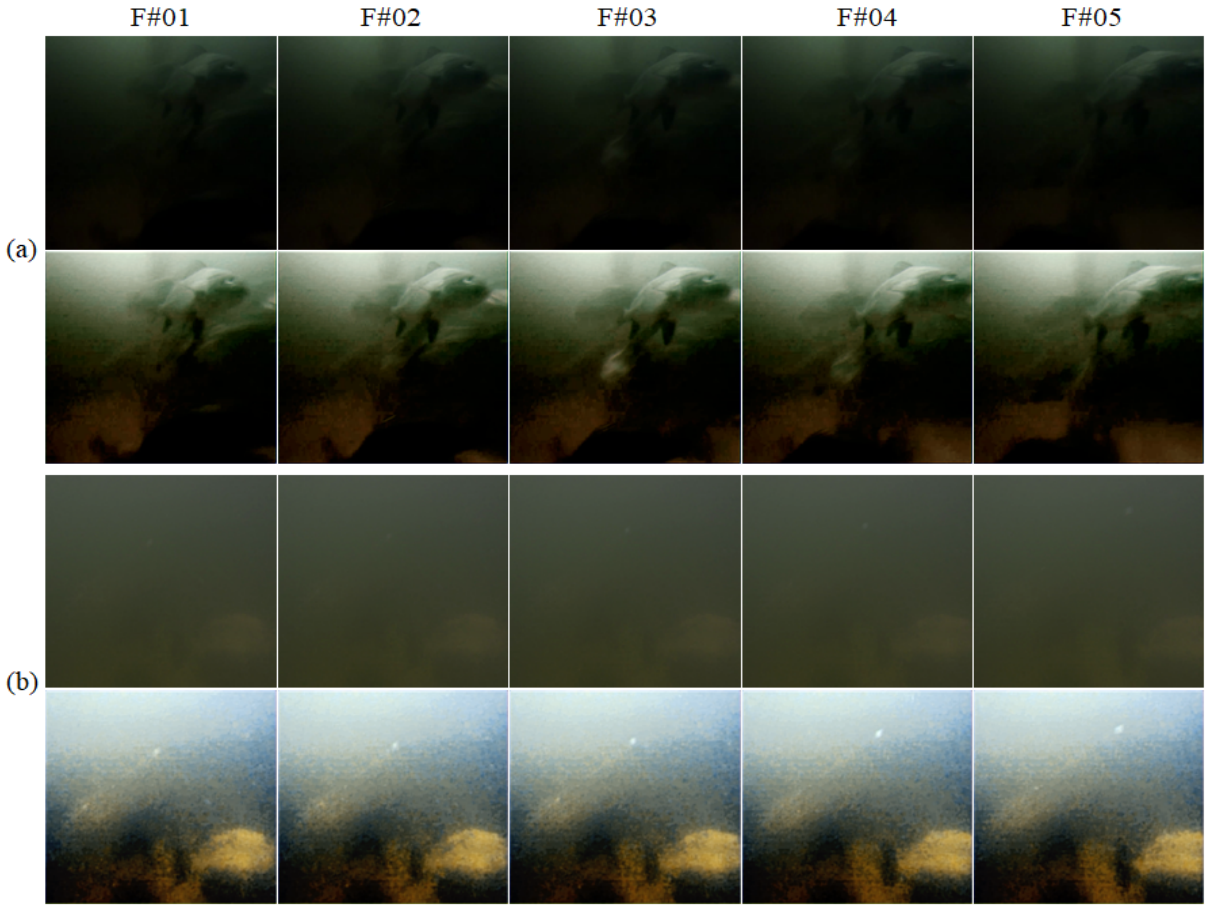


Figura 7.15: Amostras de vídeos coletados em ambientes de alta degradação com respectivas versões realçadas. (a) Vídeo: EMA_TQ1 e (b) Vídeo: EMA_TQ2.

A sequência de EMA_TQ1 foi obtida a uma profundidade de 0,5m e a de EMA_TQ2 foi coletada a uma profundidade de 0,8m, dada a baixa iluminação ambiente no interior dos tanques. Em ambas as sequências, o realce é limitado devido, principalmente, a esta reduzida

[4] `https://drive.google.com/drive/folders/1nEdm2SsHCwHs9Opu_KEYzZMrQvRGc3cS?usp=drive_link`

iluminação ambiente. Entretanto, pode-se perceber a efetividade do método na redução do efeito da turbidez e recuperação de contraste. A percepção do contorno e coloração dos animais também é facilitada nos quadros realçados.

A Fig. 7.16 mostra detalhes relativos aos *frames* #01 das sequências da Fig. 7.15, como forma de quantificar o desempenho do modelo. São indicadas a distribuição dos pixels no espaço de cores RGB, a média por canal de cor e o número de pontos discretizados (NPD) na distribuição dos pixels, onde cada ponto representa uma cor diferente. Os *frames* de entrada apresentam distribuições concentradas, indicando forte degradação. Nestes contextos, as cenas oferecem um número bastante reduzido de *features*, dificultando o realce. As versões realçadas mostram a recuperação proporcionada pelo método, com distribuições desenvolvendo maior dispersão e valores médios mais elevados. Um detalhe importante surge da comparação entre EMA_TQ1#1, que apresenta maior distribuição que EMA_TQ2#1. No entanto, suas versões realçadas mostram comportamento contrário, com a distribuição de EMA_TQ1#1 menor que de EMA_TQ2#1. Este efeito mostra a importância da luz ambiente na ação de realce. EMA_TQ2#1 apresenta melhores condições em termos de iluminação, facilitando o mapeamento de *features* e o realce.

As sequências realçadas para EMA_TQ1 e EMA_TQ2 na Fig. 7.15, mostram conteúdo de ruído, mais evidente em EMA_TQ2. Este ruído foi introduzido durante a fase de coleta dos vídeos, devido ao algoritmo de compressão de vídeo embarcado na câmera utilizada para a aquisição. No contexto de baixa luz ambiente em que os vídeos foram adquiridos, as cenas fornecem um número bastante reduzido de *features*, permitindo ao algoritmo aumentar a taxa de compressão em regiões da cena onde não são percebidas *features*. A percepção do ruído surge no vídeo descompactado e realçado.

7.2.3 Limitações identificadas ao tratamento de vídeo

Na estratégia para se alcançar consistência temporal, a substituição de 1% dos valores máximos e mínimos dos pixels pelas respectivas médias é fundamental para o bom desempenho do método. Na seção 5.3 discutiu-se a interpretação dada a condição de *outliers* que valores máximos e mínimos assumem quando da transição entre quadros sucessivos. Alterações bruscas nestes valores provocam a geração de artefatos de *flickering* no vídeo realçado.

O movimento aleatório dos objetos no vídeo subaquático está entre as principais causas de variação brusca nos valores máximos e mínimos. Em vídeos com turbidez moderada a alta, o surgimento de um objeto próximo ou muito próximo a câmera altera, instantaneamente, a percepção da distância média entre os objetos da cena e a câmera. Isto muda o perfil de realce e produz artefatos temporais no vídeo resultante. Outra causa está em alterações randômicas na refletância dos objetos entre cenas sucessivas. É o caso típico enfrentado no vídeo "Mobula", onde o deslocamento de um cardume de arraias (*Mobula munkiana*), provoca variações na refletância devido ao movimento das barbatanas, pois o animal possui distintas colorações entre as regiões dorsal e ventral, cujo movimento produz as perturbações. Estes aspectos aparentam

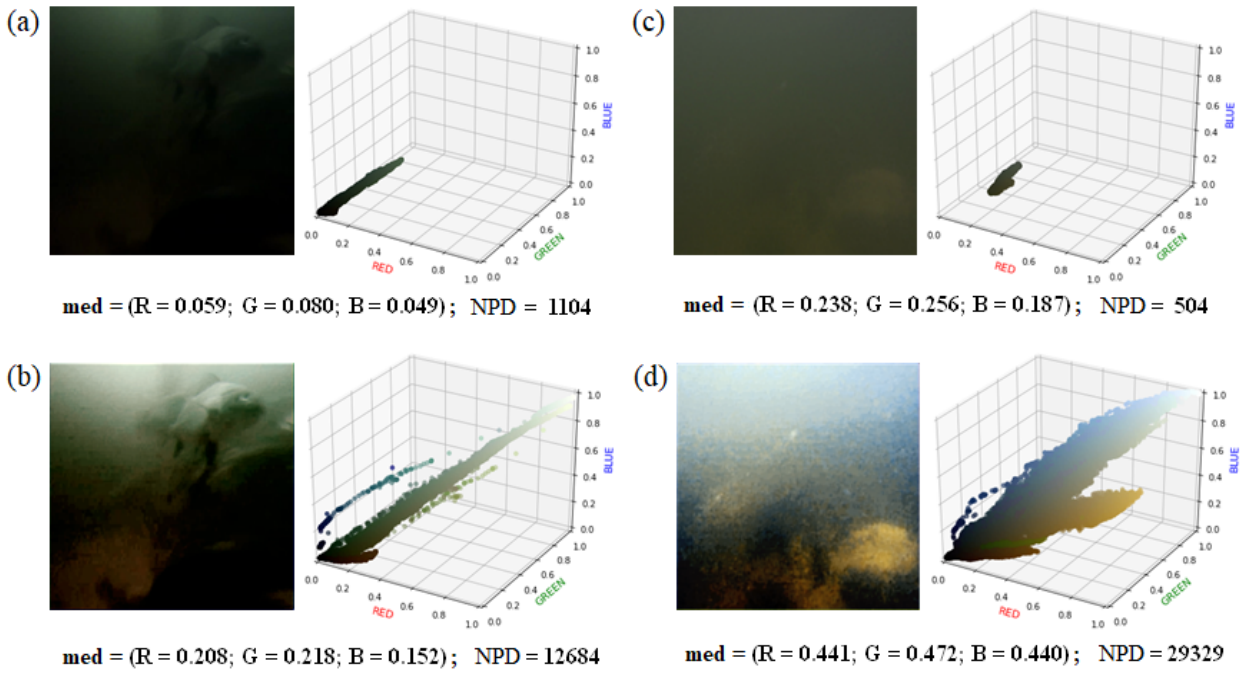


Figura 7.16: *Frames* #01 das sequências de EMA_TQ1 e EMA_TQ2 e respectivas distribuição no espaço de cores. (a) EMA_TQ1#1 e distribuição dos pixels, (b) Versão realçada de EMA_TQ1#1 e distribuição dos pixels (c) EMA_TQ2#1 e a distribuição dos pixels e (d) Versão realçada de EMA_TQ2#1 e a distribuição dos pixels.

ser específicos a determinados vídeos, mas a investigação mostra que são causas coincidentes para geração de artefatos temporais em vídeos. A intensidade destes eventos pode requer valores percentuais maiores no cálculo das médias envolvendo pixels com intensidades localizadas nas extremidades da faixa dinâmica.

Na seção 5.3, utilizou-se o percentual de 3% no estudo para a redução das variação dos parâmetros. Este percentual maior é aplicado para vídeos que apresentem variações de alta intensidade. Sequências com variações ainda mais intensas podem requerer "profundidade" superior para cálculo da média. O aumento no valor percentual produz reduções correspondentes na faixa dinâmica. Este procedimento aumenta a possibilidade de surgirem regiões saturadas na expansão do histograma e/ou de regiões escuras (preto) como resultado da contração do histograma. Outro efeito percebido no aumento do valor percentual, consiste na alteração de cor em regiões contendo pixels *outliers* da cena, afetadas pela substituição destes pixels pela média. Estas mudanças tornam-se bastante evidentes para "profundidades" elevadas no cálculo da média. Como referência, pode-se sugerir um limite confortável de 5% na profundidade para o cálculo da da média. Este valor surge de avaliações experimentais, realizadas sobre um conjunto de vídeos mais amplo que o formado pelos mencionados neste trabalho.

8 CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

A melhoria na qualidade de imagens e vídeos subaquáticos mostra grande apelo à aplicação na visão computacional. Não somente no processamento de conteúdos previamente adquiridos, mas também no importante uso em tempo real para atividades submersas. O acompanhamento remoto em tempo real de inspeções ou atividades de alto risco é demanda constante em vários áreas de pesquisa e de tecnologia aplicadas. Em anos recentes, a melhoria de imagens tem recebido grande atenção da pesquisa, com evidente evolução nas propostas e resultados. No entanto, a questão aplicação é tangencial às investigações e carece de efetividade ao interesse científico.

O presente trabalho está alinhado às perspectivas acima, ao propor metodologia aplicada a imagens e vídeos subaquáticos, mostrando forte potencial para a aplicação. A metodologia proposta usa modelo baseado em aprendizado profundo cujo treinamento considera a degradação presente na imagem, ou *frames* de vídeo, como fator que condiciona a rede a aprender a remover esta degradação.

A aplicação do método segue uma estratégia para realce no contexto subaquático, que caracteriza-se como um problema mal-posto. A informação real de cor e contraste dos objetos na cena submersa não são conhecidas na maioria dos casos. Aliado a grande diversidade de condições ambientais como turbidez, profundidade e luz ambiente, tornam esta tarefa bastante desafiadora. Estendendo a tarefa para vídeo, o desafio que se apresenta é ainda maior. Vídeos subaquáticos são muito sensíveis à variações na intensidade da ação de realce. Requer, assim, o compromisso com a consistência temporal da ação de realce.

No estágio inicial da investigação, identificou-se as principais tendências e características das propostas associadas ao realce e à restauração, onde se destacaram os modelos baseados em redes neurais. Característica fundamental destes modelos é a necessidade conjunto de dados para o treinamento das redes, onde na ampla maioria das iniciativas são requeridos dados com as respectivas informações de referência. Para o contexto subaquático, estes conjunto são muito limitados, com os autores recorrendo a sintetização destas referências. Em consonância, é bastante difundido o uso de estratégias supervisionadas para a condução do aprendizado das redes, segmentos de rede ou módulos pré-treinados ou, ainda, o pré-processamento de informações como recurso aos modelos.

O arcabouço proposto neste trabalho apresenta como contribuição uma mudança de referencial na busca pela melhoria na qualidade da imagem (*frame*). O método utiliza como fonte de informação apenas a imagem (ou vídeo) originais. Com base no conhecimento dos elementos que degradam a informação, conduz uma rede neural a aprender como removê-los da imagem, gerando no processo uma imagem com melhor percepção visual. A melhora é obtida a partir da degradação. Em outras palavras, ao invés de se conduzir o aprendizado da rede com foco em uma referência sintetizada e pré-estabelecida, parte-se da informação real dada pela

imagem original. Conduz-se o aprendizado da rede para efetivar a ação (realce) a partir do conhecimento de suas características de degradação.

O trabalho também apresenta como contribuição a proposição de uma metodologia que não requer imagens de referência, utilizando apenas a imagem original como fonte de informação. Em consequência, o modelo apresentado requer baixo nível de supervisão, com a estratégia adotada baseada em aprendizado auto-supervisionado. A remoção de degradação adicional imposta à imagem de saída se configura como tarefa pretexto para a auto-supervisão. Esta degradação adicional é coerente com a apresentada pela imagem de entrada e a rede neural aprende a compensá-la durante o treinamento.

Como contribuição secundária, o trabalho explora o conceito de imagens-guias duais para a estimativa de *features* associadas à qualidade e sua redução. A dificuldade em se ter a real percepção das cores na imagem subaquática, configuram o realce (em termos amplos, também a melhoria da qualidade) destas imagens como um problema um mal-posto. Ambas as imagens-guias são utilizadas na intensificação da degradação na versão que substitui a imagem de saída da rede na função de perda. Enquanto que em uma imagem-guia as características de degradação são acentuadas, na outra os aspectos que facilitam a percepção da cena da imagem de entrada são valorizados. Na composição da versão usada na função de perda, a imagem-guia com elementos de degradação da imagem de entrada contribui acentuando a degradação, enquanto a imagem-guia com os elementos que favorecem a qualidade é utilizada como referência para a redução da percepção destes elementos.

O método proposto foi estendido para aplicações em vídeo. Neste sentido, o trabalho apresenta como contribuição uma proposta para o tratamento de vídeo subaquático que promove, simultaneamente, melhor percepção visual e a consistência temporal. Para a obtenção da consistência temporal, o trabalho explora características específicas dos vídeos subaquáticos e associadas à variação dos seus parâmetros. Ambas as tarefas são aprendidas pelo modelo durante a etapa de treinamento.

Os resultados alcançados com a metodologia mostram desempenho bastante satisfatório do modelo desenvolvido, tanto no tratamento de imagem como vídeo. Recuperação de cor, contraste e textura, em conjunto com a redução na percepção da turbidez e *color cast* são evidentes nestes resultados. Os requisitos para consistência temporal foram atendidos na questão vídeo. A comparação realizada envolvendo propostas consolidadas e que contribuem na configuração do estado-da-arte em termos de realce, mostram o bom desempenho do arcabouço proposto.

Na finalização do presente trabalho foi possível identificar de linhas para o prosseguimento das atividades em trabalhos futuros. Embora a questão do tratamento de imagens subaquáticas visando realce e restauração mostre consistente evolução no panorama científico, a questão relativa à aplicação ainda carece de maior investigação. O contexto de vídeo mostra sinais de crescente interesse da comunidade científica como apontado por González-Sabbagh and Robles-Kelly (2023) e Hu et al. (2022). Na questão aplicação o potencial da proposta faz surgir possibilidades de estudo para adaptação e aplicação da metodologia para tratamento de vídeo em tempo-real. O atendimento dos requisitos de tempo associado à aquisição, eventual

transmissão e processamento são decisivos na efetivação de uma alternativa nesta linha. O uso na visão computacional, principalmente ligada à robótica subaquática, encontra demanda em um número expressivos de aplicações, desde a indústria de extração de óleo e gás, passando por inspeção e manutenção navais até exploração ambiental e arqueológica.

Ao longo do trabalho, em especial, durante o desenvolvimento do modelo e na avaliação preliminar de seus resultados, foi percebida a necessidade de um conjunto de dados (*dataset*) destinado ao desenvolvimento de métodos para realce e restauração. Com exceção do *dataset* Turbid, não há no âmbito científico conjunto de dados especificamente preparado para auxiliar o desenvolvimento. O Turbid foi concebido para auxiliar a avaliação de métodos e no que tange a utilização em desenvolvimento de modelos mostra limitações. Embora tenha um apelo à aplicação, a elaboração de um conjunto de dados desta natureza é bastante complexa e tem forte cunho científico. Pois, para uso no desenvolvimento são requeridos conhecimento abrangente e definições claras a cerca de cores, iluminação e estrutura da cena e adequação controlada de *features* de degradação, como turbidez e *color cast* (semelhantes ao Turbid), mas valorizando a interação destas com a luz ambiente. Assim, com base na experiência adquirida, pode-se apontar como prosseguimento, trabalhos de investigação científica e preparação para elaboração de uma base de dados para o desenvolvimento e avaliação de métodos para realce.

Em termos de investigação científica, quatro aspectos surgem da avaliação final do trabalho. O primeiro relaciona-se ao emprego das operações duais de expansão e contração do histograma. Estas operações geram as informações de referência iniciais para o método proposto baseado em degradação. Estas operações são extremamente sensíveis à variação dos parâmetros, com forte impacto na tarefa de melhoria de vídeo. A alternativa proposta para se obter a consistência temporal atua exatamente no ajuste destes parâmetros. Como sequência de estudo, sugere-se investigação para se definir uma alternativa a estas operações. Em uma concepção inicial, uma metodologia elaborada considerando a teoria Retinex parece promissora. Retinex é baseada no conceito de constância de cores e visa a obtenção de características reais da cena, eliminando os efeitos da componente de irradiação da luz ambiente sobre as cores dos objetos. Este contexto de investigação torna-se interessante no processamento de imagens e vídeos subaquáticos. Não somente pelo fato da percepção subjetiva da qualidade da imagem melhorada ser mais significativa que a avaliação quantitativa, baseada em valores numéricos oferecidos pelas métricas em uso. Mas também pela natureza do método apresentado, em que a imagem de entrada se constitui na única informação para o algoritmo e, implicitamente, baliza suas hipóteses na execução da tarefa proposta (realce ou restauração, por exemplo) relativas à constância de cores. Em um aspecto introdutório, a teoria retinex poderia ser explorada como auxílio no ajuste do resultado da expansão/contração do histograma.

O segundo aspecto é definido pela fundamentação da metodologia à luz das definições e classificação dos métodos de aprendizado auto-supervisionado. O método proposto consiste em uma abordagem auto-supervisionada, pois usa elementos dos dados de entrada na definição da tarefa-pretexto e a partir disto promove sua auto-supervisão. No entanto, não encontra posição na taxonomia associada às técnicas de auto-aprendizado considerada atualmente. Assim, torna-

se interessante um trabalho para consolidação teórica da técnica, mesclando, por exemplo, elementos da teoria do problema inverso, pois, conceitualmente, o método procura resolver a degradação como um problema inverso.

O terceiro aspecto relaciona-se a avaliação objetiva da ação de melhoria, especificamente, a questão das métricas. A partir dos procedimentos para avaliação dos resultados do método proposto, percebe-se a avaliação de imagens subaquáticas como linha de investigação envolvendo desdobramentos do trabalho. Métricas sem referência para o contexto subaquático, sugerem melhores predisposição e consistência na fundamentação para avaliação dos resultados. Isto depreende, principalmente, do fato de não se dispor de *ground truth* para imagens e, principalmente, vídeos subaquáticos, limitando o emprego de métricas referenciadas. Em adição, as métricas não-referenciadas em uso, mostram diferentes percepções sobre qualidade da imagem, que dificultam a análise e mostram incoerências com a análise subjetiva.

Finalizando, o quarto aspecto reside na perspectiva de generalização da metodologia proposta a outras áreas de estudo. A estratégia de se impor penalização adicional à função de perda durante o treinamento de um autoencoder, tem o intuito de estimular o aprendizado da rede na decodificação da informação de entrada, sem as *features* que estabelecem ou fundamentam a penalização. Isto configura um consistente panorama para investigação em outras áreas como remoção de ruído (*denoising*) em áudio, em imagens da área médica e na astronomia, recuperação de imagens subexpostas e *dehaze*. Cada área possui características específicas que determinarão como será a sua "degradação".

Referências

W. Burger, M. J. Burge, Digital Image Processing: An Algorithmic Introduction Using Java, 2nd ed., Springer Publishing Company, Incorporated, 2016.

J. Ryan, Submarine sonar scares divers, https://youtu.be/AaO6jQEmfoY, 2021.

E. P. Guerra, Mobula munkiana y aetobatus narinari, https://www.youtube.com/watch?v=csi4cw6AYB4, 2022.

S. Detloff, Los arcos puerto vallarta scuba dive, https://www.youtube.com/watch?v=eevRxlsgNts&ab_channel=Shaundetloff, 2022.

A. Saleh, M. Sheaves, D. Jerry, M. R. Azghadi, Adaptive uncertainty distribution in deep learning for unsupervised underwater image enhancement, 2022. URL: https://arxiv.org/abs/2212.08983. doi:10.48550/ARXIV.2212.08983.

M. J. Islam, Y. Xia, J. Sattar, Fast underwater image enhancement for improved visual perception, IEEE Robotics and Automation Letters 5 (2020) 3227–3234.

C. D. Mello, B. U. Moreira, P. J. D. de Oliveira Evald, P. J. L. Drews, S. S. da Costa Botelho, Underwater enhancement based on a self-learning strategy and attention mechanism for high-intensity regions, Computers & Graphics 107 (2022) 264–276.

C. Li, S. Anwar, F. Porikli, Underwater scene prior inspired deep underwater image and video enhancement, Pattern Recognition 98 (2020) 107038.

N. Ibraheem, M. Hasan, R. Z. Khan, P. Mishra, Understanding color models: A review, ARPN Journal of Science and Technology 2 (2012).

P. Drews-Jr, E. Nascimento, S. Botelho, M. Campos, Underwater depth estimation and image restoration based on single images, IEEE Computer Graphics and Applications 36 (2016) 24–35.

C. Ancuti, A. Codruta, T. Haber, P. Bekaert, Enhancing underwater images and videos by fusion, IEEE/CVF CVPR (2012) 81–88.

C. Li, S. Anwar, J. Hou, R. Cong, C. Guo, W. Ren, Underwater image enhancement via medium transmission-guided multi-color space embedding, IEEE Transactions on Image Processing 30 (2021) 4985–5000.

J. A. Donaldson, P. Drews-Jr, M. Bradley, D. L. Morgan, R. Baker, B. C. Ebner, Countering low visibility in video survey of an estuarine fish assemblage, Pacific Conservation Biology 26 (2020) 190–200.

Y. Cong, C. Gu, T. Zhang, Y. Gao, Underwater robot sensing technology: A survey, Fundamental Research 1 (2021) 337–345.

S. Fayaz, S. Parah, G. Qureshi, V. Kumar, Underwater image restoration: A state-of-the-art review, IET Image Processing 15 (2020).

N. Jerlov, Marine Optics, volume 14, 1976.

H. W. J. Rittel, M. M. Webber, Dilemmas in a general theory of planning, Policy Sciences 4 (1973).

Y. Wang, W. Song, G. Fortino, L.-Z. Qi, W. Zhang, A. Liotta, An experimental-based review of image enhancement and image restoration methods for underwater imaging, IEEE Access 7 (2019) 140233–140251.

W. Song, Y. Wang, D. Huang, D. Tjondronegoro, A rapid scene depth estimation model based on underwater light attenuation prior for underwater image restoration, Advances in Multimedia Information Processing – PCM 2018 (2018) 678–688.

S. Anwar, N. Barnes, L. Petersson, Attention prior for real image restoration, arXiv:2004.13524v1 [cs.CV] (2020).

Y. Lin, L. Shen, Z. Wang, K. Wang, X. Zhang, Attenuation coefficient guided two-stage network for underwater image restoration, IEEE Signal Processing Letters PP (2020) 1–1.

C. Y. Li, R. Mazzon, A. Cavallaro, Underwater image filtering: methods, datasets and evaluation, CoRR abs/2012.12258 (2020).

J. Y. Chiang, Y.-C. Chen, Underwater image enhancement by wavelength compensation and dehazing, IEEE Transactions on Image Processing 21 (2012) 1756–1769.

C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, P. Bekaert, Color balance and fusion for underwater image enhancement, IEEE TIP 27 (2018) 379–393.

C. Fabbri, M. J. Islam, J. Sattar, Enhancing underwater imagery using generative adversarial networks, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 7159–7165.

C. Li, C. Guo, W. Ren, R. Cong, J. Hou, S. Kwong, D. Tao, An underwater image enhancement benchmark dataset and beyond, IEEE Transactions on Image Processing 29 (2020) 4376–4389.

X. Li, G. Hou, K. Li, Z. Pan, Enhancing underwater image via adaptive color and contrast enhancement, and denoising, Engineering Applications of Artificial Intelligence 111 (2022a) 104759.

C. Li, C. Guo, L. Han, J. Jiang, M.-M. Cheng, J. Gu, C. C. Loy, Low-light image and video enhancement using deep learning: A survey, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2022b) 9396–9416.

S. Raveendran, M. Patil, G. Birajdar, Underwater image enhancement: a comprehensive review, recent trends, challenges and applications, Artificial Intelligence Review 54 (2021).

Y. Hashisho, M. Albadawi, T. Krause, U. F. von Lukas, Underwater color restoration using u-net denoising autoencoder, in: 2019 11th International Symposium on Image and Signal Processing and Analysis (ISPA), 2019, pp. 117–122.

X. Yang, H. Li, R. Chen, Underwater image enhancement with image colorfulness measure, Signal Processing: Image Communication 95 (2021) 116225.

X. Ding, Y. Wang, Y. Yan, Z. Liang, Z. Mi, X. Fu, Jointly Adversarial Network to Wavelength Compensation and Dehazing of Underwater Images, arXiv e-prints (2019) arXiv:1907.05595.

Y. Cho, H. Jang, R. Malav, G. Pandey, A. Kim, Underwater image dehazing via unpaired image-to-image translation, International Journal of Control, Automation and Systems 18 (2020) 605–614.

X. Yu, Y. Qu, M. Hong, Underwater-gan: Underwater image restoration via conditional generative adversarial network, in: Z. Zhang, D. Suter, Y. Tian, A. Branzan Albu, N. Sidère, H. Jair Escalante (Eds.), Pattern Recognition and Information Forensics, Springer International Publishing, Cham, 2019, pp. 66–75.

P. Uplavikar, Z. Wu, Z. Wang, All-in-one underwater image enhancement using domain-adversarial learning, arXiv:1905.13342v1 (2019).

A. Dudhane, P. Hambarde, P. Patil, S. Murala, Deep underwater image restoration and beyond, IEEE Signal Processing Letters 27 (2020) 675–679.

J.-Y. Zhu, T. Park, P. Isola, A. A. Efros, Unpaired image-to-image translation using cycle-consistent adversarial networks, in: 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242–2251.

Y. Li, R. Chen, Uda-net: Densely attention network for underwater image enhancement, IET Image Processing 15 (2021).

X. Zhai, A. Oliver, A. Kolesnikov, L. Beyer, S4l: Self-supervised semi-supervised learning, in: 2019 IEEE/CVF International Conference on Computer Vision (ICCV), 2019, pp. 1476–1485. doi:10.1109/ICCV.2019.00156.

K. Panetta, C. Gao, Human-visual-system-inspired underwater image quality measures, IEEE Journal of Oceanic Engineering 41 (2015) 1–11.

M. Yang, A. Sowmya, An underwater color image quality evaluation metric, IEEE Transactions on Image Processing 24 (2015) 6062–6071.

Z. Xing, M. Cai, J. Li, Improved shallow-uwnet for underwater image enhancement, in: 2022 IEEE International Conference on Unmanned Systems (ICUS), 2022, pp. 1191–1196. doi:10.1109/ICUS55513.2022.9986534.

Z. Fu, X. Lin, W. Wang, Y. Huang, X. Ding, Underwater image enhancement via learning water type desensitized representations (2021).

Y. Han, L. Huang, Z. Hong, S. Cao, Y. Zhang, J. Wang, Deep supervised residual dense network for underwater image enhancement, Sensors 21 (2021) 3289.

K. Hu, C. Weng, Y. Zhang, J. Jin, Q. Xia, An overview of underwater vision enhancement: From traditional methods to recent deep learning, Journal of Marine Science and Engineering 10 (2022) 241.

S. Zheng, Y. Ma, J. Pan, C. Lu, G. Gupta, Low-light image and video enhancement: A comprehensive survey and beyond (2022).

Q. Qi, Y. Zhang, F. Tian, Q. M. J. Wu, K. Li, X. Luan, D. Song, Underwater image co-enhancement with correlation feature matching and joint learning, IEEE Transactions on Circuits and Systems for Video Technology 32 (2022) 1133–1147.

H.-H. Chang, P.-F. Cheng, J.-K. Guo, C.-C. Sung, A self-adaptive single underwater image restoration algorithm for improving graphic quality, EURASIP Journal on Image and Video Processing 96 (2020).

M. Liu, S. Jin, C. Yao, C. Lin, Y. Zhao, Temporal consistency learning of inter-frames for video super-resolution, IEEE Transactions on Circuits and Systems for Video Technology (2022) 1–1.

M. Han, Z. Lyu, T. Qiu, M. Xu, A review on intelligence dehazing and color restoration for underwater images, IEEE Transactions on Systems, Man, and Cybernetics: Systems PP (2018) 1–13.

S. González-Sabbagh, A. Robles-Kelly, A survey on underwater computer vision, ACM Computing Surveys (2023).

B. L. McGlamery, A Computer Model For Underwater Camera Systems 0208 (1980) 221 – 231.

J. Jaffe, Computer modeling and the design of optimal underwater imaging systems, IEEE Journal of Oceanic Engineering 15 (1990) 101–111.

Y. Schechner, N. Karpel, Recovery of underwater visibility and structure by polarization analysis, IEEE Journal of Oceanic Engineering - IEEE J OCEANIC ENG 30 (2005) 570–587.

D. Berman, D. Levy, S. Avidan, T. Treibitz, Underwater single image color restoration using haze-lines and a new quantitative dataset, IEEE Transactions on Pattern Analysis and Machine Intelligence 43 (2021) 2822–2837.

D. Akkaynak, T. Treibitz, A revised underwater image formation model, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018, pp. 6723–6732. doi:10.1109/CVPR.2018.00703.

P. Baldi, Autoencoders, unsupervised learning, and deep architectures, in: I. Guyon, G. Dror, V. Lemaire, G. Taylor, D. Silver (Eds.), Proceedings of ICML Workshop on Unsupervised and Transfer Learning, volume 27 of *Proceedings of Machine Learning Research*, PMLR, Bellevue, Washington, USA, 2012, pp. 37–49. URL: https://proceedings.mlr.press/v27/baldi12a.html.

K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior, IEEE TPAMI 33 (2010) 2341–2353.

P. Drews-Jr, E. Nascimento, F. Moraes, S. Botelho, M. Campos, Transmission estimation in underwater single images, IEEE ICCVw (2013) 825–830.

D. Akkaynak, T. Treibitz, Sea-thru: A method for removing water from underwater images, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 1682–1691. doi:10.1109/CVPR.2019.00178.

K. Zuiderveld, Contrast limited adaptive histogram equalization, Graphics Gems IV (1994) 474–485.

E. Park, J.-Y. Sim, Underwater image restoration using geodesic color distance and complete image formation model, IEEE Access PP (2020) 1–1.

K. Wang, Y. Hu, J. Chen, X. Wu, X. Zhao, Y. Li, Underwater image restoration based on a parallel convolutional neural network, Remote Sensing 11 (2019) 1591.

O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: N. Navab, J. Hornegger, W. M. Wells, A. F. Frangi (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer International Publishing, Cham, 2015, pp. 234–241.

C. Desai, S. Benur, R. A. Tabib, U. Patil, U. Mudenagudi, Depthcue: Restoration of underwater images using monocular depth as a clue, in: 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), 2023, pp. 196–205. doi:10.1109/WACVW58289.2023.00025.

R. Hummel, Image enhancement by histogram transformation, Computer Graphics and Image Processing 6 (1977) 184–195.

A. Codruta, C. Ancuti, C. Vleeschouwer, P. Bekaert, Color balance and fusion for underwater image enhancement, IEEE TIP PP (2018) 1–1.

C. Liu, G. Zhong, X. Liu, J. Dong, Underwater image colour constancy based on deep sparse nonnegative matrix factorization, IET Image Processing 11 (2016).

A. S. Abdul Ghani, N. A. Mat Isa, Automatic system for improving underwater image contrast and color through recursive adaptive histogram modification, Computers and Electronics in Agriculture 141 (2017) 181–195.

D. Huang, Y. Wang, W. Song, J. Sequeira, S. Mavromatis, Shallow-water image enhancement using relative global histogram stretching based on adaptive parameter acquisition (2018) 453–465.

D. Donoho, De-noising by soft-thresholding, IEEE Transactions on Information Theory 41 (1995) 613–627.

Y. Wang, J. Zhang, Y. Cao, Z. Wang, A deep cnn method for underwater image enhancement (2017) 1382–1386.

K. He, J. Sun, X. Tang, Guided image filtering, IEEE transactions on pattern analysis and machine intelligence 35 (2013) 1397–1409.

J. Li, K. A. Skinner, R. M. Eustice, M. Johnson-Roberson, Watergan: Unsupervised generative network to enable real-time color correction of monocular underwater images, IEEE Robotics and Automation Letters 3 (2018) 387–394.

S. Anwar, C. Li, F. Porikli, Deep underwater image enhancement, CoRR abs/1807.03528 (2018).

C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, IEEE Signal processing letters 25 (2018) 323–327.

J. Li, K. A. Skinner, R. M. Eustice, M. Johnson-Roberson, WaterGAN: Unsupervised generative network to enable real-time color correction of monocular underwater images, IEEE Robotics and Automation letters 3 (2017) 387–394.

H. Zhao, O. Gallo, I. Frosio, J. Kautz, Loss functions for image restoration with neural networks, IEEE Transactions on Computational Imaging 3 (2017) 47–57.

K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition (2015).

H.-H. Yang, K.-C. Huang, W.-T. Chen, Laffnet: A lightweight adaptive feature fusion network for underwater image enhancement, in: 2021 IEEE International Conference on Robotics and Automation (ICRA), 2021, pp. 685–692. doi:10.1109/ICRA48506.2021.9561263.

J. T. Barron, A general and adaptive robust loss function (2019) 4326–4334.

R. Cong, W. Yang, W. Zhang, C. Li, C.-L. Guo, Q. Huang, S. Kwong, Pugan: Physical model-guided underwater image enhancement using gan with dual-discriminators, IEEE Transactions on Image Processing (2023) 1–1.

Y. Zhang, Q. Qi, K. Li, D. Liu, Underwater video consistent enhancement: a real-world dataset and solution with progressive quality learning, Multimedia Tools and Applications (2023) 1–27.

S.-J. Peng, H. Zhang, X. Liu, W. Fan, B. Zhong, X. Du, Real-time video dehazing via incremental transmission learning and spatial-temporally coherent regularization, Neurocomputing 458 (2020).

C. Huang, J. Li, B. Li, D. Liu, Y. Lu, Neural compression-based feature learning for video restoration, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 5872–5881. doi:10.1109/CVPR52688.2022.00578.

V. Patil, W. Van Gansbeke, D. Dai, L. Van Gool, Don't forget the past: Recurrent depth estimation from monocular video, IEEE Robotics and Automation Letters 5 (2020) 6813–6820.

H.-M. Zhang, B. Dong, A review on deep learning in medical image reconstruction, Journal of the Operations Research Society of China (2020).

S. Li, Y. Luo, Y. Zhu, X. Zhao, Y. Li, Y. Shan, Enforcing temporal consistency in video depth estimation, in: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), 2021, pp. 1145–1154. doi:10.1109/ICCVW54120.2021.00134.

H. Thimonier, J. Despois, R. Kips, M. Perrot, Learning long term style preserving blind video temporal consistency, in: 2021 IEEE International Conference on Multimedia and Expo (ICME), 2021, pp. 1–6. doi:10.1109/ICME51207.2021.9428445.

W. Gao, Y. Li, Y. Yin, M.-H. Yang, Fast video multi-style transfer, in: 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), 2020, pp. 3211–3219. doi:10.1109/WACV45572.2020.9093420.

C. Lei, Y. Xing, H. Ouyang, Q. Chen, Deep video prior for video consistency and propagation, IEEE Transactions on Pattern Analysis and Machine Intelligence (2022) 1–1.

A. Duarte, F. Codevilla, J. Gaya, S. Botelho, A dataset to evaluate underwater image restoration methods, in: OCEANS 2016, 2016, pp. 1–6.

J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database (2009) 248–255.

R. Liu, X. Fan, M. Zhu, M. Hou, Z. Luo, Real-world underwater enhancement: Challenges, benchmarks, and solutions under natural light, IEEE Transactions on Circuits and Systems for Video Technology PP (2020) 1–1.

M. J. Islam, C. Edge, Y. Xiao, P. Luo, M. Mehtaz, C. Morse, S. S. Enan, J. Sattar, Semantic segmentation of underwater imagery: Dataset and benchmark (2020) 1769–1776.

C. Li, J. Guo, C. Guo, Emerging from water: Underwater image color correction based on weakly supervised color transfer, IEEE Signal Processing Letters 25 (2018) 323–327.

J. I. Tamir, S. X. Yu, M. Lustig, Unsupervised deep basis pursuit: Learning inverse problems without ground-truth data (2020).

G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, R. Willett, Deep learning techniques for inverse problems in imaging (2020).

J. Amjad, J. Sokolić, M. R. Rodrigues, On deep learning for inverse problems (2018) 1895–1899.

A. Gómez-Ríos, S. Tabik, J. Luengo, A. Shihavuddin, F. Herrera, Coral species identification with texture or structure images using a two-level classifier based on convolutional neural networks, Knowledge-Based Systems 184 (2019) 104891.

A. Pretorius, S. Kroon, H. Kamper, Learning dynamics of linear denoising autoencoders, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, volume 80 of *Proceedings of Machine Learning Research*, 2018, pp. 4141–4150. doi:https://doi.org/10.48550/arXiv.1806.05413.

A. Odena, V. Dumoulin, C. Olah, Deconvolution and checkerboard artifacts, Distill 1 (2016).

D. Charte, F. Charte, S. García, M. J. Del Jesus, F. Herrera, A practical tutorial on autoencoders for nonlinear feature fusion: Taxonomy, models, software and guidelines, Information Fusion 44 (2018) 78–96.

C. Aggarwal, Neural Networks and Deep Learning: A Textbook, Springer International Publishing, 2018. URL: https://books.google.com.br/books?id=AsTswQEACAAJ. doi:https://doi.org/10.1007/978-3-319-94463-0.

C. Liu, H. Li, S. Wang, M. Zhu, D. Wang, X. Fan, Z. Wang, A dataset and benchmark of underwater object detection for robot picking, IEEE Int. Conf. on Multimedia Expo Workshops (ICMEW). (2021).

M. Pedersen, J. Haurum, R. Gade, T. Moeslund, N. Madsen, Detection of marine animals in a new underwater dataset with varying visibility, 2019.

S. Raine, R. Marchant, P. Moghadam, F. Maire, B. Kettle, B. Kusy, Multi-species seagrass detection and classification from underwater images, 2020.

B. Boom, P. Huang, J. He, R. Fisher, Supporting ground-truth annotation of image datasets using clustering, 2012, pp. 1542–1545.

M. Yuval, I. Alonso, G. Eyal, Y. Loya, A. Murillo, T. Treibitz, Repeatable semantic reef-mapping through photogrammetry and label-augmentation, Remote Sensing 13 (2021) 659.

L. Li, E. Rigall, J. Dong, G. Chen, Mas3k: An open dataset for marine animal segmentation (2021) 194–212.

P. Zwilgmeyer, M. Yip, A. Teigen, R. Mester, A. Stahl, The varos synthetic underwater data set: Towards realistic multi-sensor underwater data with ground truth. (2021) 3715–3723.

M. Ferrera, V. Creuze, J. Moras, P. Trouvé, Aqualoc: An underwater dataset for visual-inertial-pressure localization., The International Journal of Robotics Research 38 (2019) 027836491988334.

D. P. Kingma, J. Ba, Adam: A method for stochastic optimization (2017).

N. Khare, P. S. Thakur, P. Khanna, A. Ojha, Analysis of loss functions for image reconstruction using convolutional autoencoder., 2022, pp. 338–349. doi:10.1007/978-3-031-11349-9-30.

Y. Wang, N. Li, Z. Li, Z. Gu, H. Zheng, B. Zheng, M. Sun, An imaging-inspired no-reference underwater color image quality assessment metric, Computers & Electrical Engineering 70 (2018) 904–913.

Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE TIP 13 (2004) 600–612.

W. Xue, L. Zhang, X. Mou, A. Bovik, Gradient magnitude similarity deviation: A highly efficient perceptual image quality index, IEEE TIP 23 (2013).

G. Sharma, W. Wu, E. Dalal, The CIEDE2000 color-difference formula: Implementation notes, supplementary test data, and mathematical observations, Color Research & Application 30 (2005) 21 – 30.

L. Zhang, L. Zhang, X. Mou, FSIM: A feature similarity index for image quality assessment, IEEE TIP 20 (2011) 2378 – 2386.

E. L. N. Cams, Octopus hide and seek, https://www.youtube.com/watch?v=kIGJ7RtKeYo, 2018.

A. Rojas, 13 de febrero de 2023, https://www.youtube.com/watch?v=4kEdT3ey4Mg&ab_channel=ArturoRosas, 2023.

K. Ernst, Saint martin, https:https://www.youtube.com/watch?v=RemoOfLIix8&list=PLfp9wAed7xH5IdNDjIhgvO0XOU_uJW3In&index=12&ab_channel=KenErnst, 2022.

D. Trecker, Monitoring fish farm - underwater rov [uncut], https://www.youtube.com/watch?v=Wm4Js_ZKrM0&ab_channel=DeepTrekkerInc, 2021.

H. Wu, L. Liao, C. Chen, J. Hou, A. Wang, W. Sun, Q. Yan, W. Lin, Disentangling aesthetic and technical effects for video quality assessment of user generated content (2022).

A. Mittal, R. Soundararajan, A. C. Bovik, Making a "completely blind" image quality analyzer, IEEE Signal Processing Letters 20 (2013) 209–212.

M. A. Saad, A. C. Bovik, C. Charrier, Blind prediction of natural video quality, IEEE Transactions on Image Processing 23 (2014) 1352–1365.