CAST-GAN: LEARNING TO REMOVE COLOUR CAST FROM UNDERWATER IMAGES

Chau Yi Li and Andrea Cavallaro

Centre for Intelligent Sensing, Queen Mary University of London

ABSTRACT

Underwater images are degraded by blur and colour cast caused by the attenuation of light in water. To remove the colour cast with neural networks, images of the scene taken under white illumination are needed as reference for training, but are generally unavailable. As an alternative, one can use surrogate reference images taken close to the water surface or degraded images synthesised from reference datasets. However, the former still suffer from colour cast and the latter generally have limited colour diversity. To address these problems, we exploit open data and typical colour distributions of objects to create a synthetic image dataset that reflects degradations naturally occurring in underwater photography. We use this dataset to train Cast-GAN, a Generative Adversarial Network whose loss function includes terms that eliminate artefacts that are typical of underwater images enhanced with neural networks. We compare the enhancement results of Cast-GAN with four state-of-the-art methods and validate the cast removal with a subjective evaluation.

Index Terms— Image enhancement, Underwater images, Generative Adversarial Networks, Image synthesis

1. INTRODUCTION

Generative Adversarial Networks (GANs) [1] are successfully used in image enhancement tasks such as colour retouching [2], deblurring [3] and colourisation [4]. The success of GANs is made possible by the availability of training images, such as expert-retouched reference images [5] or artificially degraded images [3, 4].

An important image enhancement problem is removing the colour cast from underwater images, whose typical blue or green cast is caused by the selectively attenuated illuminant in water. The ambient light attenuated along the vertical depth results in the varying water colour observed at different depths. Objects are illuminated by this attenuated light and the reflected lights are scattered and absorbed, thus reducing its intensity along the scene-to-camera distance (range) [6, 7]. Hence objects farther from the camera appear blurred and lack contrast. Moreover, the light scattered towards the camera veils the object with the water colour.

Underwater image enhancement can use neural networks [8–15] or physics-based models [9, 16–19]. *Neural networks* for underwater image enhancement are trained using images taken closer to the water surface, which however are still affected by some colour cast [8]. Alternatively, training images are synthesised with neural networks [14, 15] or from reference images with physics-based models [9–12]. However, these methods may not consistently model the physics of the degradation, which increases with the range [9–11], or the veiling water colour [9, 10, 12], thus leading to inconsistent enhancements. Moreover, while the range of the scene is available in some datasets [10, 12], such as NYU-D [20] and Middlebury [21], these datasets depict indoor scenes and have limited colour variation. *Physics-based methods* either do not account for the light attenuation along the vertical depth, thus the enhanced images are still under some colour cast [16–18, 22], or remove the cast with global white balancing, thus distorting the colour of water regions [19].

In this paper, we propose a neural network-based enhancement method, Cast-GAN, which uses the trained generator of a GAN to remove the colour cast from underwater images, without distorting the colour of water regions. We tailor the loss function of the generator to the underwater image enhancement problem by encouraging the enhancement of edges and colours towards the radiance under a white illuminant. Moreover, using a physics-based model [23], we synthesise a reference training dataset that reflects the colour degradations in different water types. The resulting dataset is unique in its ability to consistently account for scattering effects and light attenuation along the depth. Cast-GAN successfully enhances images with heavy colour casts. We compare Cast-GAN against four methods that are based on physics-based models or neural networks. Finally, we validate the cast removal by Cast-GAN with a subjective experiment, which shows a preference for Cast-GAN-enhanced images as more likely to be taken under white illuminant.

2. PROPOSED METHOD

In this section, we discuss the synthesis of the training images using a physics-based degradation process. We also propose a loss function that specifically corrects distortions introduced by an underwater image enhancement process.

2.1. Underwater image degradation

Let $I(x, y)$ be an underwater image and $J(x, y)$ be the radiance of the same scene under a white illumination, i.e. without the presence of water. The illuminant on objects at depth D is the ambient light $E(x, y) = e^{-\kappa D}$, where κ is the diffuse attenuation coefficient [24]. Pure water mass, which does not contain reflective objects, has the colour of the ambient light. The colour-casted radiance, $E(x, y)J(x, y)$, is further absorbed and scattered along the range, $r(x, y)$. The loss in intensity can be described as an exponential decay modulated by the beam attenuation coefficient, β, as [6]:

$$t(x, y) = e^{-\beta r(x,y)}. \quad (1)$$

Scattering redirects the propagation path and increases, along the range, the spatial spread of the reflected light, thus blurring details of objects farther from the camera. The spatial spread can be derived from the knowledge of the range and water properties [25]. A portion of ambient light, that is backscattered along the range, increases the water colour cast on objects with $\frac{\alpha}{\beta}E(x, y)$ [6], where α is the scattering coefficient.

The degraded radiance captured in $I(x, y)$ is the combination of the attenuated radiance and backscattered ambient light:

$$I(x, y) = t(x, y)S\big(E(x, y)J(x, y)\big) + \big(1 - t(x, y)\big)\frac{\alpha}{\beta}E(x, y), \quad (2)$$

where $S(\cdot)$ denotes the scattering process. Pure water mass can be represented by the ambient light intensity $E(x, y)$.

To create training images that reflect underwater degradations, we need the radiance under white illuminant, $J(x, y)$, the scene's range, $r(x, y)$, and the water properties, which can be obtained from empirical data [24, 26].

2.2. Reference images

We create the reference images of radiance under white illuminant, $J(x, y)$, by modifying information provided in the SINTEL dataset [27]. We use the clean pass images that account for light interacting with and occluded by objects in the scene, generating specular reflections, shading, and cast shadows, which are added to the original albedo [27]. The dataset includes the object segmentation masks and their ranges in full resolution, $r'(x, y)$, stored as the Z-buffer (Fig. 1).

The limited colour palette in SINTEL does not reflect the colour distribution in natural images. We recolour the images to ensure that the overall colour distribution in $J(x, y)$ follows that of natural images and also the physics of light interactions. To this end, we sample the colour distribution from the Berkeley Segmentation Dataset [28] of each object segmentation mask to avoid bias towards large areas such as sky and grass. We fit a Gaussian Mixture Model (GMM) to the RGB distribution of each segmentation mask, using the number of components that maximises the ratio of variances between and within components [29]. We use the ensemble of all individual GMMs as the overall colour distribution for $J(x, y)$ and assign an RGB value sampled from the ensembled GMM to each segmentation in SINTEL images. We preserve the light interactions by maintaining the lightness values in CIELab colour space in SINTEL [30].

Furthermore, we create pure water regions by assigning the most faraway objects with $J(x, y) = E(x, y)$. Specifically, we assign regions with $r'(x, y) \geq r_\tau$ as pure water regions, where r_τ is chosen to determine the extent of the water region(s) in the image. We choose the water region(s) to be 10% of the synthesised image by selecting r_τ as the 90$^{\text{th}}$ percentile of all the ranges in the image.

When synthesising the degraded image $I(x, y)$, we ensure some radiances from objects in $J(x, y)$ remain after attenuation. However, most objects in SINTEL have large $r'(x, y)$ values (over 20m), the distance at which, because of attenuation, most colours will appear black in water. Instead of using $r'(x, y)$ directly in the synthesis, we use the relative ranges between objects by normalising $r'(x, y)$ to $[r_{\text{min}}, r_{\text{max}}]$, whose values determine the attenuation extent observed in the image. We obtain the range of objects (with $r'(x, y) < r_\tau$) used in the synthesis as

$$r(x, y) = r_{\text{min}} + \frac{r'(x, y) - \min r'(x, y)}{r_\tau - \min r'(x, y)}(r_{\text{max}} - r_{\text{min}}). \quad (3)$$

The portion of intensity reaching the camera after attenuation, $t(x, y)$, is hence

$$t(x, y) = \begin{cases} e^{-\beta r(x,y)} & \text{if} \quad r'(x, y) < r_\tau \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

From the above, we obtain the reference images with unattenuated radiance under white illuminant and their intensity loss due to attenuation. In the next section, we describe how we implement the range-dependent scattering that blurs the image.

2.3. Degraded images

We model the range-dependent scattering with Gaussian blurring, where the spatial spread, w, of the scattered light depends on the filter's variance [25]. We successively apply an additive Gaussian filter, $g(x, y)$ [31], of variance σ_0^2. The cumulative variance, σ^2, after k applications is then $k\sigma_0^2$. For simplicity, we only describe the scattering process $S(\cdot)$ on $J(x, y)$.

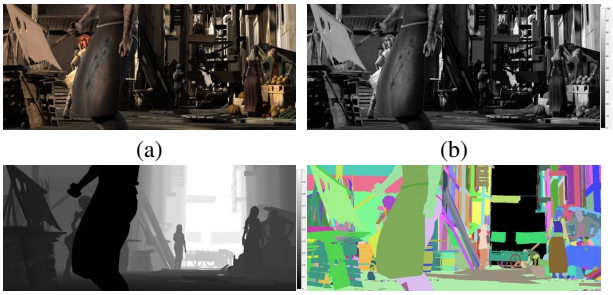

Fig. 1. For the synthesis of the training dataset, we use open data provided in SINTEL [27]. (a) Sample SINTEL image that accounts for the light interactions between objects in the scene. As the original limited colour palette is unsuitable for direct synthesis of training images, we resample the colour distribution from the Berkeley Segmentation Dataset [28] and recolour objects. (b) Lightness channel in CIELab colour space [30] that reflects the light interactions. (c) Scene's range normalised to $[0, 1]$. Faraway regions assigned as pure water are shown in white. (d) Segmentation masks for objects. Each colour represents an object with the same reflectance. The synthesised degraded images are shown in Fig. 3.

The spatial spread w of the scattered light along the range around the pixel position (x', y') can be expressed as $w^2 \propto r^3(x', y')$ [25]. We also aim to ensure the spread is proportional to the cumulative variance after successive Gaussian blurring, i.e. $w^2 \propto \sigma^2$. To this end, we decompose the scene, based on range, into N layers. The farthest objects are in the N^{th} layer and will be the most blurred. We define the $k^{\text{th}} \in \{1, 2, ..., N\}$ layer, of the same dimension as $J(x, y)$, as

$$L_k(x, y) = \begin{cases} J(x, y) & \text{if } \left(\frac{k-1}{N}\right)^{\frac{2}{3}} < \frac{r(x,y)}{r_{\text{max}}} \leq \left(\frac{k}{N}\right)^{\frac{2}{3}} \\ 0 & \text{otherwise.} \end{cases} \quad (5)$$

The process is initialised with $S_N(x, y) = L_N(x, y) * g(x, y)$, where $*$ is the convolution operator. From $k = N - 1$, we aggregate the unattenuated layer $L_k(x, y)$ and the already blurred layers, and apply Gaussian blurring as

$$S_k(x, y) = (L_k(x, y) + S_{k+1}(x, y)) * g(x, y). \quad (6)$$

The scattered light is obtained as $S(J(x, y)) = S_1(x, y)$, where the radiance in $S_k(x, y)$ is effectively blurred k times. Fig. 2 shows an example of the successive blurring process. When synthesising the degraded images (Eq. 2), we avoid abrupt intensity changes by also blurring $t(x, y)$ with the same $g(x, y)$ and layer decomposition (Eq. 5).

We synthesise the degraded training images for 10 Jerlov water types, using κ from [24] and α, β from [26], from the reference images obtained in Sec. 2.2. To synthesise a dataset with diverse illuminant conditions caused by attenuation along the depth, we randomly sample D from $[0.2\ , D_{\text{max}}]$ for each synthesised image, where D_{max} ensures the illuminant $E(x, y) \geq 0.05$ for all colour channels in the water type. The minimum D of 0.2 ensures the attenuation along the depth is visible in the image. For the range, $r_{\text{min}} = 0.2$ is the minimum range at which an object is in focus, whereas $r_{\text{max}} = 20$ is the maximum range before most colours appear black. Fig. 3 shows example images from our synthesised training dataset, which contains 690 images for each water type.

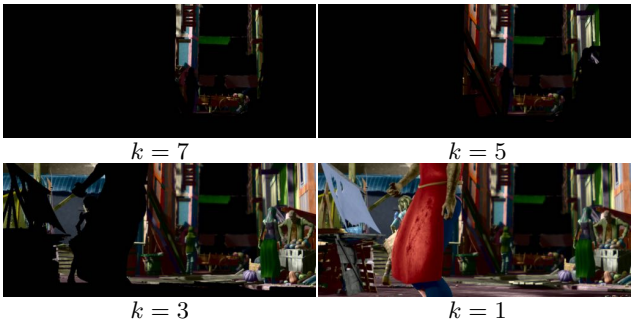

Fig. 2. The range-dependent scattering blurs scene details and increases the spatial spread of the reflected light, with successive applications of a Gaussian blur filter. The unattenuated scene is decomposed, based on range, into $N = 7$ non-overlapping layers. Objects farthest from the camera are in the N^{th} layer. Starting from step $k = 7$, we blur the aggregation of the k^{th} layer and layers blurred in previous steps. Note that the appearance of an object is increasingly blurred at each step. At the last step, the objects farthest away from the camera (present in $k = 7$) are most the blurred, whereas object that are only present at $k = 1$ remain sharp.

2.4. Loss function

We aim to train the generator $G(\cdot)$ of Cast-GAN to enhance a degraded image under colour cast, $I(x, y)$, to produce an enhanced radiance $J^*(x, y) = G(I(x, y))$ that is similar to the reference unattenuated radiance, $J(x, y)$, under white illumination and without scattering. The similarity is determined by the loss function. The discriminator learns to distinguish between $J^*(x, y)$ and $J(x, y)$. However, the typical cross-entropy, $\mathcal{L}_{\text{GAN}}$, used to encourage the generator to mislead the discriminator [1], produces images with checkerboard and droplet artefacts [32]. Fig. 4(a) shows an example of GAN enhanced image with only $\mathcal{L}_{\text{GAN}}$.

The loss function we propose for Cast-GAN, $\mathcal{L}_{\text{Cast-GAN}}$, encourages the similarity to reference images by combining four terms:

$$\mathcal{L}_{\text{Cast-GAN}} = \lambda_{\text{GAN}}\mathcal{L}_{\text{GAN}} + \lambda_{\text{L}_1}\mathcal{L}_{\text{L}_1} + \lambda_{\text{MSSIM}}\mathcal{L}_{\text{MSSIM}} + \lambda_{\text{perc}}\mathcal{L}_{\text{perc}}, \quad (7)$$

where the hyper-parameters $\lambda_{\{\cdot\}}$ control the balance between the amount of enhancement and image quality. While these terms have been used in other enhancement tasks [4,32], we are the first to combine them for underwater image enhancement.

$\mathcal{L}_{\text{L}_1}$ measures the intensity difference between the enhanced and reference image [4] as $\mathcal{L}_{\text{L}_1} = \|J^*(x, y) - J(x, y)\|_1$, where $\|\cdot\|_1$ is the absolute value. $\mathcal{L}_{\text{MSSIM}}$, a multi-scale structural similarity, measures the quality of the reconstruction, in terms of luminance, contrast and structure between $J^*(x, y)$ and $J(x, y)$ at M scales [33]. The luminance comparison, $l(\cdot, \cdot)$, is only at the original resolution. At each scale m, $J^*(x, y)$ and $J(x, y)$ are downsampled to half dimensions of the previous scale, starting from the original resolution ($m = 1$), denoted by J^*_m and J_m, respectively. If we denote the contrast comparison and the structure comparison as $c_m(\cdot, \cdot)$ and $s_m(\cdot, \cdot)$, respectively, the multi-scale structural similarity is defined as [33]

$$\mathcal{L}_{\text{MSSIM}} = \Big[l(J^*_1, J_1)\Big]^a \prod_{m=1}^{M} \Big[c_m(J^*_m, J_m)\Big]^{p_m} \Big[s_m(J^*_m, J_m)\Big]^{q_m}, \quad (8)$$

where a, p_m and q_m weighs each comparison. The larger the weight, the more important the component is to the similarity measure.

Finally, we control the image quality explicitly with the so-called perceptual loss, $\mathcal{L}_{\text{perc}}$, between enhanced and reference image [32]. $\mathcal{L}_{\text{perc}}$ uses the intermediate layers of a neural network, $\Phi(\cdot)$, to extract image features. The more similar the features are, the more perceptually similar the images are. The perceptual loss also helps to suppress checkerboard artefacts caused by the cross-entropy $\mathcal{L}_{\text{GAN}}$ [34] and is defined as [32]

$$\mathcal{L}_{\text{perc}} = \|\Phi(J^*(x, y)) - \Phi(J(x, y))\|_2^2, \quad (9)$$

where $\|\cdot\|_2$ is the Euclidean distance.

Fig. 3. A synthesised image for each of the 10 Jerlov water types [24] for a scene at depth 1.30m under the water surface, and with the farthest object at 1.65m from the camera. (a) Oceanic water (top to bottom: type I, IA, IB, II, III). Red is the most attenuated, inducing a blue cast. (b) Coastal water (type 1, 3, 5, 7, 9). Blue is the most attenuated, inducing a green cast.

3. VALIDATION

We validate the proposed loss function with an ablation study and compare Cast-GAN with four state-of-the-art methods: two neural networks, namely U-Net Denoising (U-Net) [8] and Underwater Scene Prior Inspired Image Enhancement (UWCN) [12], and two physics-based methods, namely Depth-dependent Background Light (DBL) [16] and Underwater Haze Line (UWHL) [19]. UWHL and Cast-GAN are the only methods explicitly aiming to remove colour cast. The trained networks and codes of all methods are available online, except U-Net, whose results were provided by the authors.

The generator of Cast-GAN consists of 6 ResNet blocks [35], with parameters initialised by normal distribution to prevent the vanishing gradient problem. The discriminator is a 70×70 PatchGAN [4] that determines the similarity between patches in the enhanced and reference images. The networks are trained with a learning rate of 1×10^{-5} for 150 epochs, which are then exponentially decreased to 0 in 150 epochs. In all epochs, to prevent overfitting, only a 436×436 patch is randomly selected from each 1024×436 training image. For the loss function, we use as $\Phi(\cdot)$ the third ReLU layer of the VGG-16 network [36]. $\mathcal{L}_{\text{MSSIM}}$ uses $M = 5$ scales, with $p = q =$ [0.04, 0.29, 0.30, 0.24, 0.13]. We select as hyper-parameters of the loss function $\lambda_{\text{GAN}} =1$, $\lambda_{\text{L}_1} =100$, $\lambda_{\text{MSSIM}} =100$ and $\lambda_{\text{perc}} =10$. The large differences in the hyper-parameters suggest $\mathcal{L}_{\text{GAN}}$ stabilises over epochs and requires a larger contributions from the remaining terms to mitigate the artefacts introduced by $\mathcal{L}_{\text{GAN}}$. Cast-GAN was trained for 13 hours on an NVIDIA Tesla V100 with CUDA 9.0 using the ADAM optimiser.

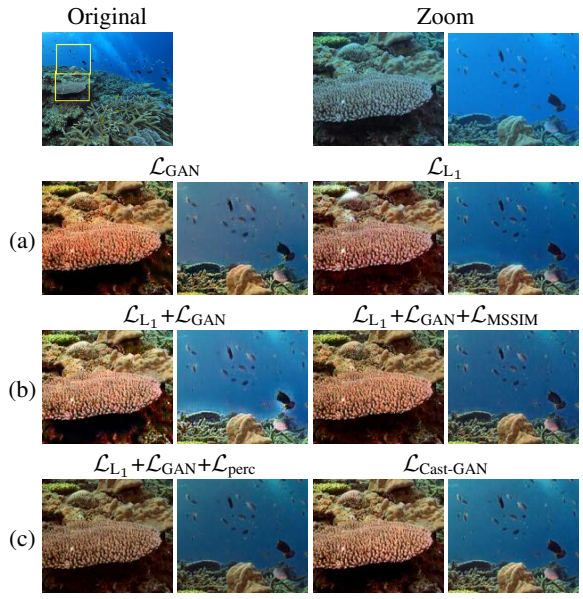

Fig. 4. The proposed loss function, $\mathcal{L}_{\text{Cast-GAN}}$, mitigates the artefacts of individual components and improves visual quality. (a) Artefacts include checkerboard effects caused by $\mathcal{L}_{GAN}$ and unnatural white regions caused by $\mathcal{L}_{L_1}$. (b) Using $\mathcal{L}_{MSSIM}$ enhances the edges on the coral. (c) Using $\mathcal{L}_{perc}$ removes checkerboard effects. Note that we show the areas identified by yellow rectangles in the original image.

Fig. 4 shows sample results from the ablation study on the proposed loss function, $\mathcal{L}_{\text{Cast-GAN}}$. We tested individually the typical $\mathcal{L}_{GAN}$ for GANs and $\mathcal{L}_{L_1}$ for colour intensity similarity. Using only $\mathcal{L}_{GAN}$ introduces checkerboard artefacts that reduce the image quality, whereas using only $\mathcal{L}_{L_1}$ introduces unnaturally bright regions (Fig. 4(a)). Combining the two reduces the bright regions but results in more visible checkerboard effects (Fig. 4(b)). Adding $\mathcal{L}_{perc}$ to the combination significantly reduces checkerboard effects, whereas adding $\mathcal{L}_{MSSIM}$ enhances the edges on the corals (Fig. 4(c)). $\mathcal{L}_{\text{Cast-GAN}}$, combining the four components, enhances the colour and the edge details without introducing noticeable artefacts.

Fig. 5 shows sample enhanced images for the methods under comparisons. Most methods successfully enhance images with mild colour casts, except UWCN that adds a grey-pink cast to the scene (Fig. 5(a)). Methods that only address the degradation along the range have limited performance under heavy casts, showing little improvement on colour (Fig. 5(b,c)). Although UWHL can remove the blue colour cast, it overcorrects the colour of the objects to an unnaturally red appearance. The visibility of objects in images taken in coastal waters with heavy green casts is very limited (Fig. 5(d,e)). Both UWHL and Cast-GAN are able to remove the heavy cast and to reveal details in the scene. In particular, Cast-GAN enhances the appearance of the scuba diver in the background. Cast-GAN, however, may reduce details in some regions, such as in the darken regions in Fig. 5(c).

To validate the colour cast removal ability of Cast-GAN, we asked participants to select, in a double stimulus test, the image with objects that look more likely to be pictured under white illumination. Pairs of images included Cast-GAN against the original, UWHL [19] against the original, and UWHL [19] against Cast-GAN. Participants could select either of the images or 'cannot decide'. We collected a total of 751 responses from 21 participants on 12 original images. Table 1 shows that, when compared against the original image, Cast-GAN is selected by the participants 60.2% of the times. UWHL is

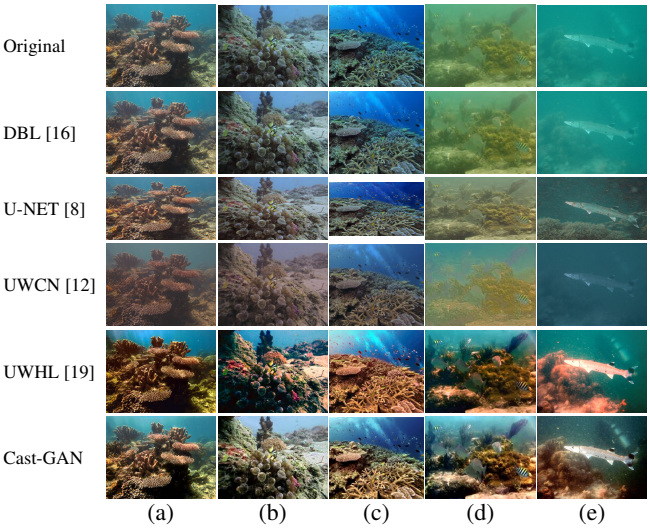

Fig. 5. Sample enhanced images taken in (a)(b)(c) oceanic water with blue casts, and (d)(e) coastal waters with heavy green casts. Note how Cast-GAN removes the cast and enhances details, while maintaining realistic colours, especially in images with heavy casts. Since U-Net can only process images of specific resolutions, we select here the image provided by authors, whose resolution is the closest to the original image's aspect ratio. For the purpose of visualisation, images in the same column may have been resized to have the same width.

Table 1. Results of the subjective experiment. Participants were asked to select from a pair of images the one whose objects appeared more likely to have been pictured under white illuminant, or 'cannot decide'. The higher the preference, the better the cast removal. Cast-GAN is consistently selected over the original image and UWHL [19].

Image pair		**Preference** (%)		
Method 1	**Method 2**	**Cannot decide**	**Method 1**	**Method 2**
Original	UWHL [19]	7.7	55.8	36.4
Original	Cast-GAN	6.0	33.8	60.2
UWHL [19]	Cast-GAN	9.8	18.7	71.5

only selected 36.4% of the times, whereas the original image is selected 55.8% of the times. When compared against UWHL, Cast-GAN is preferred to UWHL 71.5% of the times. To summarise, Cast-GAN is consistently preferred.

The complete image enhancement results can be found at `http://cis.eecs.qmul.ac.uk/projects/Cast-GAN`.

4. CONCLUSION

We proposed Cast-GAN, a Generative Adversarial Network to remove the colour cast from underwater images. We overcome the lack of training datasets by creating reference images that incorporate the effects of scattering and attenuation along the depth. We showed that our synthesised reference dataset, which accounts for the complete underwater image degradation effects, can improve the performance of enhancement networks. We assessed the effectiveness of the proposed loss function with an ablation study and validated the cast removal performance with a subjective evaluation, which shows the consistent performance of Cast-GAN over state-of-the-art methods. Cast-GAN can be extended to address other physics-based image degradations, such as haze, for which reference training images are unavailable.

5. REFERENCES

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. Int. Conf. Neural Info. Process. Syst.*, Dec. 2014, pp. 2672–2680.

[2] J. Park, J. Lee, D. Yoo, and I. Kweon, "Distort-and-recover: Color enhancement using deep reinforcement learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2018, pp. 5928–5936.

[3] S. Nah, T. H. Kim, and K. M. Lee, "Deep multi-scale convolutional neural network for dynamic scene deblurring," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, July 2017, pp. 257–265.

[4] P. Isola, J. Zhu, T. Zhou, and A. A Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2017.

[5] V. Bychkovsky, S. Paris, E. Chan, and F. Durand, "Learning photographic global tonal adjustment with a database of input / output image pairs," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2011, pp. 97–104.

[6] Y. Y. Schechner and N. Karpel, "Clear underwater vision," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2004, pp. 536–543.

[7] S. Q. Duntley, "Light in the sea," *J. Opt. Soc. Am.*, vol. 53, no. 2, pp. 214–233, Feb. 1963.

[8] Y. Hashisho, M. Albadawi, T. Krause, and U. Freiherr von Lukas, "Underwater color restoration using U-Net denoising autoencoder," in *Proc. Int. Symp. on Image Signal Process. Anal.*, Sep. 2019, pp. 117–122.

[9] Y. Shin, Y. Cho, G. Pandey, and A. Kim, "Estimation of ambient light and transmission map with common convolutional architecture," in *OCEANS 2016 MTS/IEEE Monterey*, Sep. 2016, pp. 1–7.

[10] Y. Hu, K. Wang, X. Zhao, H. Wang, and Y. Li, "Underwater image restoration based on convolutional neural network," in *Proc. Asian Conf. on Mach. Learning*, Nov. 2018, vol. 95, pp. 296–311.

[11] Y. Wang, J. Zhang, Y. Cao, and Z. Wang, "A deep CNN method for underwater image enhancement," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2017, pp. 1382–1386.

[12] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, pp. 1–11, Feb. 2020.

[13] J. Perez, A. C. Attanasio, N. Nechyporenko, and Pedro J. Sanz, "A deep learning approach for underwater image enhancement," in *Proc. Biomedical Applications Based on Natural and Artificial Computing*, June 2017, pp. 183–192.

[14] M. J. Islam, Y. Xia, and J. Sattar, "Fast underwater image enhancement for improved visual perception," *IEEE Robot. Automat. Letters*, vol. 5, no. 2, pp. 3227–3234, 2020.

[15] C. Fabbri, M. J. Islam, and J. Sattar, "Enhancing underwater imagery using generative adversarial networks," in *IEEE Int. Conf. Robot. Automat.*, May 2018, pp. 7159–7165.

[16] C. Y. Li and A. Cavallaro, "Background light estimation for depth-dependent underwater image restoration," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2018, pp. 1528–1532.

[17] A. Galdran, D. Pardo, A. Picon, and A. Alvarez-Gila, "Automatic Red-Channel underwater image restoration," *J. Vis. Commun. Image Represent.*, vol. 26, pp. 132–145, Jan. 2015.

[18] S. Emberton, L. Chittka, and A. Cavallaro, "Underwater image and video dehazing with pure haze region segmentation," *Comput. Vis. Image Underst.*, vol. 168, pp. 145–156, Mar. 2018.

[19] D. Berman, T. Treibitz, and S. Avidan, "Diving into haze-lines: Color restoration of underwater images," in *Proc. British Mach. Vis. Conf.*, Sep. 2017, pp. 44.1–44.12.

[20] N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor," in *Proc. Int. Conf. Comput. Vis. Workshop*, Nov. 2011, pp. 601–608.

[21] D. Scharstein, R. Szeliski, and R. Zabih, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," in *Proc. IEEE Workshop on Stereo and Multi-Baseline Vis.*, Dec 2001, pp. 131–140.

[22] Y. T. Peng and P. C. Cosman, "Underwater image restoration based on image blurriness and light absorption," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1579–1594, Apr. 2017.

[23] B. L. McGlamery, "A computer model for underwater camera systems," in *Ocean Opt. VI, SPIE*, Mar. 1980, vol. 0208.

[24] N. G. Jerlov, *Marine optics*, Elsevier, Jan. 1976.

[25] S. Premoze, M. Ashikhmin, R. Ramamoorthi, and S. Nayar, "Practical rendering of multiple scattering effects in participating media," in *Proc. of Eurographics Workshop on Rendering Techniques*, June 2004, pp. 363–373.

[26] M. G. Solonenko and C. D. Mobley, "Inherent optical properties of Jerlov water types," *Appl. Opt.*, vol. 54, no. 17, pp. 5392–5401, June 2015.

[27] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in *Proc. European Conf. on Comput. Vis.*, Oct. 2012, pp. 611–625.

[28] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Int. Conf. Comput. Vis.*, July 2001, vol. 2, pp. 416–423.

[29] T. Caliński and J Harabasz, "A dendrite method for cluster analysis," *Commun. in Statist.*, vol. 3, no. 1, pp. 1–27, 1974.

[30] K. McLaren, "XIII: The development of the CIE 1976 (L*a*b*) uniform colour space and colour-difference formula," *J. Soc. Dyers Colour*, vol. 92, no. 9, pp. 338–341, Sep. 1976.

[31] H. Nguyen, *GPU Gems 3*, Addison-Wesley Professional, Aug. 2007.

[32] J. Johnson, A. Alahi, and F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *European Conf. on Comput. Vis.*, Oct. 2016, pp. 694–711.

[33] Z. Wang, E. Simoncelli, and A. Bovik, "Multiscale structural similarity for image quality assessment," in *Asilomar Conf. on Signals, Syst. and Comput.*, Dec. 2003, vol. 2, pp. 1398 – 1402.

[34] A. Odena, V. Dumoulin, and C. Olah, "Deconvolution and checkerboard artifacts," *Distill*, Oct. 2016.

[35] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognit.," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, June 2016, pp. 770–778.

[36] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Int. Conf. on Learning Represent.*, May 2015.