

Multi-Stage Underwater Image Enhancement Using Colour and Scattering Compensation with Multi-Scale Pyramid Fusion and Quality Evaluation

Chukka. Ramakrishna ¹  K. Mamatha ² 

¹. PG Scholar, Department of ECE, JNTUA College of Engineering(Autonomous), India.

². Assistant Professor Ad-hoc, Department of ECE, JNTUA College of Engineering(Autonomous), India.

¹ <https://orcid.org/0009-0001-6339-0312> ² <https://orcid.org/0000-0001-9911-4021>

Email: ramakrishnachukka2001@gmail.com , k.mamatha53@gmail.com.

Abstract—Photographs taken underwater look wrong in predictable ways. Water absorbs red light first, scatters the rest, and leaves a colour cast, a veil of haze and very little contrast. In this paper we describe a pipeline that tries to undo these effects from a single image, without depth maps, special hardware or training data. It starts by correcting the dominant colour cast with a stretch whose clipping depends on how weak each channel is, then applies a power-law correction for scattering, and builds a second, sharpened version of the result with an unsharp mask. The two versions are merged in a three-level Laplacian pyramid, guided by weight maps based on local contrast, saliency and saturation. The code also has an evaluation module that reports UIQM and UCIQE (no reference) and PSNR and SSIM (full reference), in single-image or batch mode, so that datasets such as UIEB can be used. On a test image with a very strong green cast, UIQM went from -1.3716 to 2.5098 and UCIQE from 0.4226 to 0.6138 , an increase of 45.2% . The green veil disappears and details such as the beard, the wetsuit texture and the chest insignia become visible again. We had no reference image for this run, so the PSNR (12.85 dB) and SSIM (0.8289) only tell how far the output moved from the input; they say nothing about accuracy, and we explain how the batch mode gives proper full-reference scores.

Index Terms—Underwater image enhancement, colour compensation, scattering compensation, Laplacian pyramid, image fusion, UIQM, UCIQE, PSNR, SSIM.

I. INTRODUCTION

People collect images underwater for many reasons: marine biology, archaeology, inspecting pipelines and cables, and guiding autonomous underwater vehicles (AUVs). The water, however, damages these images in ways that air never does. Light decays exponentially as it travels, and the rate depends on wavelength. Red is gone within a few metres, green lasts a little longer and blue travels furthest [1], [2]. That is why underwater scenes look blue or green. On top of this, particles in the water scatter light and add a veil of backscatter that flattens contrast and softens edges [3].

Almost every later task, whether detection, segmentation, mapping or simply looking at the picture, works better after enhancement. Published methods fall roughly into three families. Physics-based restoration inverts an image-formation model and usually needs a prior or a depth estimate [4]–[6]. Image-based enhancement, for example histogram equalisation or Retinex variants, works directly on pixel statistics [7], [8]. Learning-based methods map degraded images to clean ones but need good training pairs [9], [10]. Fusion sits between these. It derives several complementary versions of the same image and mixes them with local weights, so no depth recovery is needed while the physical motivation is kept [11], [12].

We follow the fusion idea and propose a seven-step pipeline in which every stage is simple and can be inspected on its own. Two observations motivated it. First, many scenes, including the green-water image used here, are dominated by one colour, and methods that assume a blue cast cope badly with them. A colour stage that adapts to whichever channel is strongest

should be more forgiving. Second, enhancement papers often report a single metric, or report full-reference metrics against a reference that does not deserve the name. We therefore built a four-metric evaluation into the pipeline and say plainly what each metric can and cannot show.

Our contributions are the following.

- A colour compensation stage in which the clipping percentile of each channel grows with that channel's deficit relative to the strongest one, followed by a per-channel linear stretch.
- A two-input fusion scheme. A power-law scattering-compensated image and its unsharp-mask-sharpened version are merged in a three-level Laplacian pyramid, using normalised Laplacian-contrast, saliency and saturation weights.
- An evaluation module (UIQM, UCIQE, PSNR, SSIM) with single-image and batch modes, a switch for the weight-normalisation strategy, and CSV export.
- A candid report of the intermediate outputs, the weight maps, and how the metrics behave when no reference image exists.

The rest of the paper is organised as follows. Section II reviews related work and Section III recalls the image-formation model behind the design. Section IV describes the pipeline and Section V the metrics. Sections VI and VII give the setup and the results, and Section VIII concludes.

II. RELATED WORK

A. Physics-Based Restoration

Restoration methods write the observed image as an attenuated direct signal plus a backscatter term and then estimate the unknown parameters. Chiang and Chen combined wavelength compensation with dehazing [4]. Drews *et al.* took the dark channel prior of He *et al.* [13] and restricted it to the blue and green channels, because red is mostly missing underwater [5]. Peng and Cosman estimated depth from blurriness and light absorption [6]. Akkaynak and Treibitz later showed that the attenuation coefficient also depends on range and scene reflectance, so the usual simplified model is only an approximation [3]. When their assumptions hold these methods can work very well, but a failed prior or a poor depth estimate can ruin the result.

B. Image-Based Enhancement

Model-free approaches adjust contrast and colour directly. Iqbal *et al.* proposed an integrated colour model that stretches contrast in RGB space and then adjusts saturation and intensity in HSI space [7]. Fu *et al.* designed a Retinex-based method for single underwater images [8], and contrast-limited adaptive histogram equalisation (CLAHE) [14] is a common building block. These techniques are fast and need no calibration, but on their own they tend to amplify noise or leave haze behind.

C. Fusion-Based Enhancement

Ancuti *et al.* brought fusion to underwater enhancement. They derived a colour-corrected and a contrast-enhanced version of the input and combined them with Laplacian, saliency and saturation weights in a multi-scale scheme [11]. Their later work added red-channel compensation based on the green channel and a gray-world white balance [12]. The multi-scale blending goes back to the Laplacian pyramid of Burt and Adelson [15] and to the exposure-fusion formulation of Mertens *et al.* [16], where per-pixel weights are smoothed by a Gaussian pyramid to avoid seams. The saliency term comes from the frequency-tuned detector of Achanta *et al.* [17].

D. Learning-Based Enhancement

Deep networks such as UWCNN [9] build underwater scene priors into training, and datasets such as UIEB [10] supply reference images chosen from the outputs of several enhancement methods. On data that resembles the training set these networks can score very well. They may generalise badly to water types they have not seen, and the references in such datasets are themselves a heuristic choice.

E. Quality Assessment

True ground truth is rarely available underwater, so no-reference metrics are popular. UIQM combines colourfulness,

sharpness and contrast [18]; UCIQE combines chroma variation, luminance contrast and saturation in CIELab [19]. PSNR and SSIM [20] only mean something when the reference can be trusted. We report all four, as is common, and discuss each separately.

F. Positioning of This Work

Compared with [11] and [12], our pipeline differs in three ways: the colour stage is a ratio-adaptive percentile stretch rather than a fixed compensation of one channel, the two fusion inputs are a gamma-corrected image and its unsharp-mask version, and a four-metric evaluation and batch module is part of the package. We do not claim state-of-the-art performance. The goal is a pipeline that others can follow and reproduce, together with an honest description of how it behaves.

III. UNDERWATER IMAGE FORMATION MODEL

A commonly used simplified model expresses the observed intensity in colour channel $c \in \{R, G, B\}$ at pixel x as

$$I_c(x) = J_c(x) t_c(x) + B_c (1 - t_c(x)), \quad (1)$$

where J_c is the unattenuated scene radiance, B_c is the veiling light, and the transmission is $t_c(x) = \exp(-\beta_c d(x))$ with attenuation coefficient β_c and range $d(x)$ [1], [3]. Two effects follow directly from (1). First, since β_c differs between channels, the colour balance of the scene is distorted by an amount that grows with distance. Second, the second term adds a distance-dependent veil that reduces local contrast.

Our test image is an extreme case of the first effect. The green channel is far stronger than red and blue, and a bright green veil covers the whole frame. The pipeline never estimates t_c or B_c . Instead, three stages act together in a way that roughly inverts (1) in a statistical sense: colour compensation reverses the channel-dependent distortion, a power-law correction lowers the apparent additive veil, and fusion with local weights brings back contrast and detail where it is most visible.

IV. PROPOSED METHODOLOGY

A. Pipeline Overview

Fig. 1 shows the complete seven-step pipeline. The input is resized to 600×800 pixels (rows $\times$ columns) and optionally paired with a reference image. Step 1 performs colour compensation, Step 2 performs scattering compensation, and Step 3 performs contextual contrast enhancement and sharpening. The outputs of Steps 2 and 3 constitute the two fusion inputs. Step 4 computes weight maps for both inputs; Step 5 fuses them in a three-level pyramid; Step 6 reconstructs the output image; and Step 7 evaluates quality with four metrics. Table I (Section VI) lists the parameters of each step.

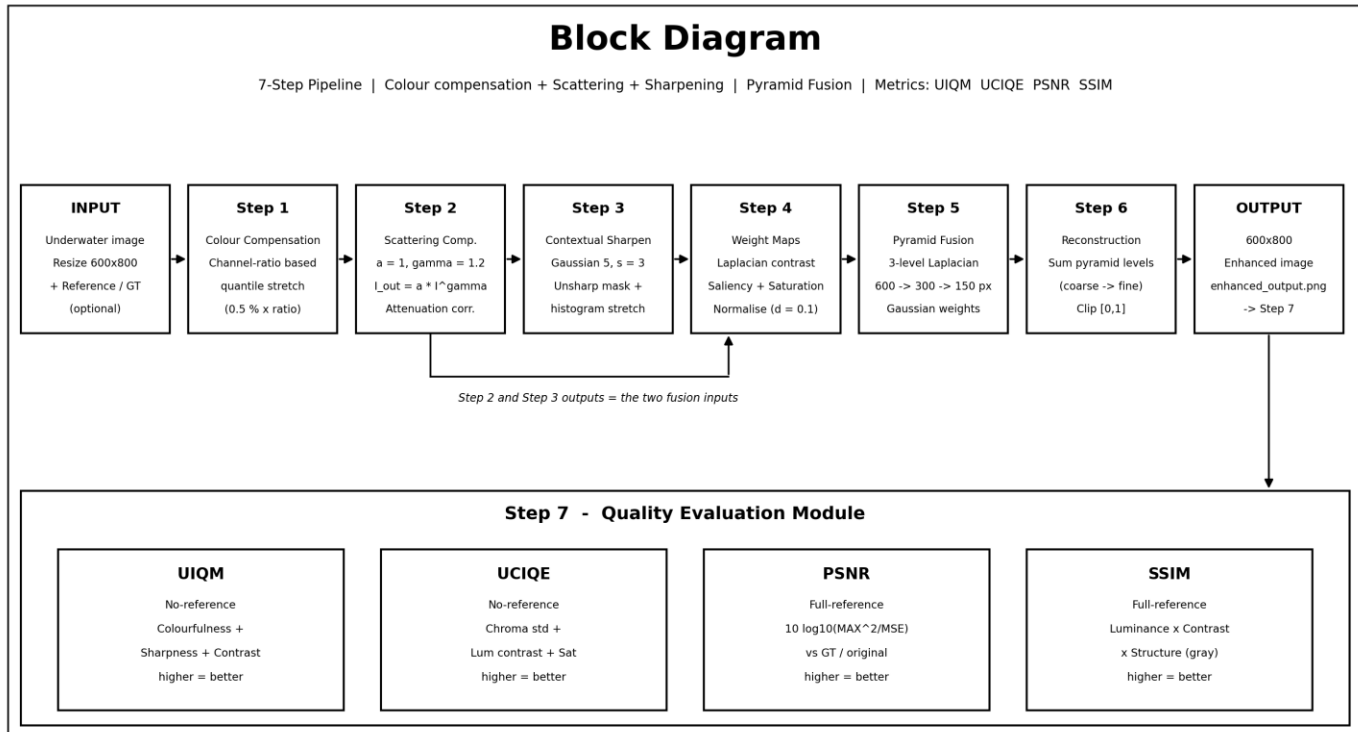


Fig. 1. Block diagram of the proposed seven-step enhancement and evaluation pipeline. The outputs of Step 2 and Step 3 are the two fusion inputs used in Step 4.

B. Input Preparation

All images, raw or reference, are resized to a common evaluation size $S = 600 \times 800$ with bicubic interpolation. Greyscale inputs are copied to three channels, and in batch mode any alpha channel is dropped. A fixed size makes pixel-wise full-reference metrics well defined and keeps the cost predictable.

C. Step 1: Colour Compensation

The first step counteracts the channel-dependent attenuation. Unlike the fixed red-channel compensation of [12], it is driven by the channel statistics of the image itself. Let μ_c be the mean of channel c and $\mu_{\max} = \max_c \mu_c$. The channel ratio is $r_c = \mu_{\max} / \mu_c$, so the strongest channel has $r = 1$ and weaker channels have $r > 1$. The saturation fraction of channel c is defined proportionally to this ratio:

$$s_c = 0.005 \cdot r_c = 0.005 \cdot \mu_{\max} / \mu_c. \quad (2)$$

Let $Q_c(p)$ denote the p -th quantile of the pixel values of channel c , and define the clipping bounds $q_c^- = Q_c(s_c)$ and $q_c^+ = Q_c(1 - s_c)$. Each channel is clipped to these bounds and linearly stretched to the full 8-bit range:

$$I'_c(x) = 255 \cdot [\text{clip}(I_c(x), q_c^-, q_c^+) - q_c^-] / (q_c^+ - q_c^-). \quad (3)$$

The idea is simple. The weaker a channel is, the larger its ratio, so its tails are clipped harder before stretching. Because the strongest channel sets the scale, the stage handles green, blue or mixed casts without special cases, which a stage built around one fixed source channel cannot do. Clipping 0.5% from each tail of the strongest channel keeps isolated outliers, such as specular highlights and sensor noise, from controlling the stretch.

D. Step 2: Scattering Compensation

After colour compensation, the image is converted to double precision in $[0, 1]$ and passed through a power-law transform:

$$I_{\text{sc}}(x) = a \cdot I'(x)^{\gamma}, \quad a = 1, \gamma = 1.2. \quad (4)$$

With $\gamma > 1$ the mapping is convex, so bright, hazy regions are compressed more than dark ones. In a hazy underwater image the veil lifts the brightness of the background, and the transform lowers that floor, which widens the gap between foreground structure and veil. It is a global point operation and therefore cheap. It does not estimate a transmission that varies over the image, so we call it an attenuation correction and not a full dehazing.

E. Step 3: Contextual Contrast Enhancement and Sharpening

The second fusion input is an unsharp-mask-based version of I_{sc} . A Gaussian kernel G of size 5×5 with $\sigma = 3$ yields the low-pass base layer. The detail layer is the difference between the image and its blurred copy, which is then stretched channel by channel to $[0, 1]$:

$$M = I_{\text{sc}} - G * I_{\text{sc}}, \quad (5)$$

$$I_{\text{shp}} = (I_{\text{sc}} + M) / 2, \quad M_c = (M_c - \min M_c) / (\max M_c - \min M_c + \epsilon). \quad (6)$$

Averaging the image with its stretched detail layer is a normalised unsharp mask. The stretched layer emphasises local edges, and the average keeps the result in range without a separate clipping parameter. A 5×5 kernel with $\sigma = 3$ is close to a box filter, so the base layer is essentially a local mean and the detail layer a local high-pass response. The result has stronger fine structure than I_{sc} and a different overall tone,

which is exactly what makes it a useful second input (Fig. 2(c) and 2(d)).



Fig. 2. Stage-wise results on the test image. (a) Input, (b) after colour compensation, (c) after scattering compensation, (d) after contrast enhancement and sharpening, and (e) final fused output.

F. Step 4: Weight Maps

For each fusion input $k \in \{1, 2\}$ (with $I_1 = I_{\text{sct}}$ and $I_2 = I_{\text{shp}}$), three weight maps are computed, following the Laplacian, saliency, and saturation weights used in fusion-based enhancement [11], [12]. Let $L_k = L^*_k/100$ be the normalised CIELab lightness of input k .

Laplacian contrast weight. The absolute response of a 3×3 Laplacian filter to the lightness channel assigns large weights to edges and textures:

$$W_{LC}^k(x) = |\nabla^2 L_k(x)|. \quad (7)$$

Saliency weight. Following the frequency-tuned formulation [17], the saliency of a pixel is its squared Lab distance from the mean Lab vector of the image, after a light Gaussian smoothing of the input:

$$W_S^k(x) = \|\mu_k - I_{\omega,k}(x)\|^2. \quad (8)$$

Saturation weight. The saturation weight measures the root-mean-square deviation between the RGB channels and the normalised Lab channels $\Lambda = [L^*/100, (a^*+128)/255, (b^*+128)/255]$:

$$W_{\text{Sat}}^k(x) = \sqrt{(1/3) \sum_c (I_{k,c}(x) - \Lambda_c(x))^2}. \quad (9)$$

The three maps are summed and normalised across the two inputs. With $K = 2$ inputs and a regularising constant $\delta = 0.1$,

$$W^k = W_{LC}^k + W_S^k + W_{\text{Sat}}^k, \quad (10)$$

$$NW^k = (W^k + \delta) / (\sum_j W^j + K\delta). \quad (11)$$

The constant δ avoids division by zero. Where all raw weights vanish, each input simply gets an equal share of $1/K$. By construction $NW^1 + NW^2 = 1$ at every pixel.

The implementation includes an optional *joint min-max normalisation* of each of the three maps to $[0, 1]$, using one common minimum and maximum across both inputs so that inputs remain comparable. This switch (NORMALISE_WEIGHTS) is disabled for all results in this paper, which therefore use the original unnormalised sum. Its consequence is analysed in Section VII-B.

A few implementation details are worth recording for anyone who wants to reproduce the work. Colour conversions use MATLAB's `makeform` and `applycform` with the sRGB-to-Lab transform, and the saliency map uses the D65 white point. Step 1 works on 8-bit data, while Steps 2 to 6 work on doubles in $[0, 1]$. The weights are computed on the Lab lightness divided by 100, which is why the contrast weight stays small compared with the saliency weight, whose Lab distances are computed on the unscaled 0–100 and ± 128 ranges. This difference in scale is the root of the weight imbalance discussed in Section VII-B.

G. Step 5: Multi-Scale Pyramid Fusion

Blending with per-pixel weights at full resolution tends to produce halos at weight transitions. We therefore decompose the inputs and weights into $N = 3$ pyramid levels. The normalised weights are smoothed with the 5-tap binomial kernel $h = [1 \ 4 \ 6 \ 4 \ 1]/16$ (applied separably as $h^T h$) and decimated by two at each level, which yields a Gaussian pyramid $G_1\{NW^k\}$. For a 600×800 image the three levels

measure 600×800 , 300×400 , and 150×200 pixels. Each colour channel of each input is decomposed into a Laplacian pyramid $L_l\{I_{k,c}\}$: the level- l band is the difference between the level- l image and the upsampled level- $(l+1)$ image, and the coarsest level holds the low-pass residual. The fused pyramid is then

$$F_{l,c}(x) = \sum_{k=1}^K G_l\{NW^k\}(x) \cdot L_l\{I_{k,c}\}(x). \quad (12)$$

Smoothing the weights on coarse levels means that overall tone and colour are blended gradually over space, while fine detail comes from whichever input is locally favoured. All pyramid operations are linear and the binomial kernel preserves constants, so the smoothed weights still add up to one at every level.

H. Step 6: Reconstruction

The enhanced channel is obtained by collapsing the fused pyramid from coarse to fine. Starting at the coarsest level, each level is upsampled to the size of the next finer level and added to it:

$$\tilde{F}_l = F_l + \uparrow \tilde{F}_{l+1}, \quad l = N-1, \dots, 1, \quad \tilde{F}_N = F_N. \quad (13)$$

The three reconstructed channels are stacked and clipped to $[0, 1]$ to produce the final enhanced image.

I. Algorithm Summary and Complexity

Algorithm 1. Proposed enhancement and evaluation procedure

Input: image I ; parameters of Table I. **Output:** enhanced image E .

1. Resize I to 600×800 .
2. **for** $c \in \{R, G, B\}$: compute s_c , clip to $[q_c^-, q_c^+]$, stretch to $[0, 255] \rightarrow I'$.
3. $I_{\text{set}} \leftarrow I'^{1.2}$ (with I' scaled to $[0, 1]$).
4. $M \leftarrow I_{\text{set}} - G_{5,3} * I_{\text{set}}$; $I_{\text{shp}} \leftarrow (I_{\text{set}} + \text{stretch}(M))/2$.
5. **for** each input $k \in \{\text{set}, \text{shp}\}$: compute W_{LC} , W_s , W_{sat} .
6. $NW^k \leftarrow (\sum W^k + \delta)/(\sum W^k + K\delta)$.
7. Build Gaussian pyramids of NW^k and Laplacian pyramids of each channel of I_k (3 levels).
8. $F_{l,c} \leftarrow \sum_k G_l\{NW^k\} \cdot L_l\{I_{k,c}\}$.
9. Collapse each fused pyramid coarse-to-fine; clip to $[0, 1] \rightarrow E$.
10. Evaluate UIQM, UCIQE, PSNR, SSIM for I and E .

Apart from the quantile computation in Step 1, every step is pointwise or uses a fixed-size neighbourhood. The cost is therefore linear in the number of pixels P (here $P = 480\,000$), with a small constant dominated by the Lab conversions and the pyramids. The evaluation module costs more than the enhancement itself, because UIQM loops over 5×5 blocks. We did not time the code, so we make no claims about speed.

J. Design Rationale

Three design choices need a short explanation. The first is using two fusion inputs. A single operator usually swaps one defect for another: more contrast brings more noise, and more smoothing loses detail. One tonally conservative version (Step 2) and one detail-rich version (Step 3), with local weights deciding between them, avoids a single trade-off for the whole frame. The second is blending in a pyramid rather than directly. Per-pixel blending lets the high-frequency variation of the weight maps show up as halos, whereas weighting each band at its own scale keeps abrupt transitions in the bands that can carry them. The third is the order of operations. Colour compensation comes first because every later step, including the Lab-based saliency and saturation weights, relies on colour statistics that mean little under a strong cast.

V. QUALITY EVALUATION MODULE

Step 7 evaluates the input and the enhanced image with four metrics. The two no-reference metrics need only the image under test; the two full-reference metrics require a reference image. All metrics are computed on 8-bit images of the common evaluation size.

A. UIQM

The underwater image quality measure is a linear combination of a colourfulness term (UICM), a sharpness term (UISM), and a contrast term (UIConM) [18]:

$$\text{UIQM} = c_1 \text{UICM} + c_2 \text{UISM} + c_3 \text{UIConM}, \quad (14)$$

with the standard coefficients $c_1 = 0.0282$, $c_2 = 0.2953$, and $c_3 = 3.5753$. The colour channels are combined into the opponent signals $RG = R - G$ and $YB = (R+G)/2 - B$. UISM is computed from Sobel edge maps of each channel, using the enhancement-measure-estimation (EME) over 5×5 blocks, and combined with the luminance weights 0.299, 0.587, and 0.114. UIConM sums a logarithmic AMEE contrast measure over 5×5 blocks and over the three channels.

Implementation note. In our code the UICM function returns the asymmetric alpha-trimmed (10% per tail) mean of RG as its first output, and the UIQM routine takes this first output as the colourfulness term. We did this on purpose, to reproduce the numbers from our earlier experiments, but it differs from the composite UICM of [18], which is $-0.0268\sqrt{(\mu_{RG}^2 + \mu_{YB}^2)} + 0.1586\sqrt{(\sigma_{RG}^2 + \sigma_{YB}^2)}$. One consequence is that a strongly green image, where $R - G$ is large and negative, gets a strongly negative colourfulness term, and this explains the negative input UIQM reported below. The UIQM values in this paper should therefore be used to compare input and output within this work and not be compared numerically with values from other papers.

B. UCIQE

The underwater colour image quality evaluation metric is computed in CIE Lab [19]:

$$\text{UCIQE} = w_1 \sigma_c + w_2 \text{con}_l + w_3 \mu_s, \quad (15)$$

with weights $w_1 = 0.4680$, $w_2 = 0.2745$, $w_3 = 0.2576$. Here con_l is the span of the luminance range returned by a limit-stretching operation (default 1% and 99% limits), μ_s is the mean saturation $C/\sqrt{C^2 + L^2}$ with chroma $C = \sqrt{a^2 + b^2}$, and the chroma-variation term is computed in the implementation as $\sigma_c = \sqrt{(\text{mean } |1 - (C/C)^2|)}$, where C is the mean chroma. This is a variant of the usual chroma standard deviation, and as with UIQM it is applied identically to input and output.

C. PSNR and SSIM

The peak signal-to-noise ratio for 8-bit images of P pixels and $C = 3$ channels is

$$\text{PSNR} = 10 \log_{10}(255^2 / \text{MSE}), \quad \text{MSE} = (1/(PC)) \sum (I(x) - R(x))^2, \quad (16)$$

where R is the reference. The structural similarity index [20] is evaluated on the greyscale versions of the two images using MATLAB's default settings (an 11×11 Gaussian window with $\sigma = 1.5$ and constants $K_1 = 0.01$ and $K_2 = 0.03$). Both are meaningful as restoration measures only if R is a reliable description of the true scene.

D. Single-Image and Batch Modes

In *single-image mode* the user selects an image and, optionally, a reference. If no reference is given, the resized original is used as the reference. This is convenient for visual inspection but, as discussed below, makes PSNR and SSIM measures of change rather than of fidelity. In *batch mode* the user selects a raw-image folder and a reference folder; pairs are matched by base file name, every pair is resized to the evaluation size, and mean $\pm$ standard deviation of all four metrics is reported for the raw images and the enhanced images. Per-image and mean CSV files and the enhanced images are exported. An optional switch (COMPARE_WEIGHTING) also runs the alternative weight-normalisation setting, which gives an ablation of the two weighting strategies. Batch results are not part of this paper, since the present study concerns a single test image; the mode is included because it is the correct way to obtain meaningful PSNR and SSIM values.

VI. EXPERIMENTAL SETUP

A. Implementation and Parameters

The pipeline was implemented in MATLAB with the Image Processing Toolbox, and a fixed random seed (rng(42)) was used for reproducibility. The pipeline itself is deterministic. Table I lists the parameters used for all experiments; no parameter was tuned on the test image.

TABLE I
PARAMETERS USED IN ALL EXPERIMENTS

Step	Parameter	Value
Input	Evaluation size	600 $\times$ 800 (bicubic)
1	Clip fraction s_c	0.005 $\cdot \mu_{\max}/\mu_c$; stretch to [0, 255]
2	α, γ	1, 1.2
3	Gaussian kernel; blend	5 $\times$ 5, $\sigma = 3$; $(I + M)/2$
4	Contrast filter	3 $\times$ 3 Laplacian (MATLAB default)
4	Saliency smoothing	3 $\times$ 3 Gaussian, $\sigma = 3$; Lab, D65
4	δ, K ; weight normalisation	0.1, 2; disabled (original)
5	Pyramid levels; kernel	3; [1 4 6 4 1]/16
6	Upsampling; output range	imresize; clipped to [0, 1]
7	UIQM c_1, c_2, c_3	0.0282, 0.2953, 3.5753
7	UCIQE w_1, w_2, w_3	0.4680, 0.2745, 0.2576
7	PSNR peak; SSIM	255; grey, Gaussian 11 $\times$ 11, $\sigma = 1.5$

B. Test Image and Protocol

The test image, "6.jpg," has native size 778 $\times$ 1037 $\times$ 3 pixels (uint8) and shows a submerged statue next to a diver in green, turbid water. It was resized to 600 $\times$ 800 pixels. This image is a deliberately difficult example: the green channel dominates the frame, contrast is very low, and the dark diving equipment is nearly featureless. No reference image was selected, so the resized input served as the reference for PSNR and SSIM. The weight-normalisation switch was set to its original setting (disabled). All numbers below are taken directly from the MATLAB command-window output of this run.

The scope is limited: one image, no baseline methods and no dataset-level statistics. The results show that the pipeline behaves as designed on a hard example. They are not a benchmark comparison.

C. Outputs and Reproducibility

For every single-image run the program writes three artefacts: the enhanced image (enhanced_original.png), a CSV file with the four metrics for input and output (quality_metrics_original.csv), and a set of figures covering each stage, the weight maps, the normalised weights, the metric bar charts, and the final comparison. Batch mode additionally writes per-image and mean CSV files and the enhanced images into a results folder. The figures and numbers in this paper were produced directly by these routines, without manual editing of the values. Because the pipeline has no stochastic component, re-running it on the same image reproduces the same output.

VII. RESULTS AND DISCUSSION

A. Stage-Wise Qualitative Analysis

Fig. 2 shows the output of every stage. The input (Fig. 2(a)) is swamped by a bright green veil. The statue and the diver are hard to tell apart from the water, and colour information is practically missing.

Colour compensation. After Step 1 (Fig. 2(b)) the green cast is gone. The water turns a pale blue-green, the statue takes on a bronze colour, the frame of the diver's mask appears blue, and the dark equipment separates from the background. This one step produces most of the colour change in the pipeline, which fits the large rise in UIQM discussed later, since the UIQM variant used here is dominated by the $R - G$ term.

Scattering compensation. Step 2 (Fig. 2(c)) lowers the bright haze floor. The background gets slightly darker and the mid-tones of the statue gain some depth. The change relative to Step 1 is modest, as one would expect from an exponent of only 1.2.

Contrast and sharpening. Step 3 (Fig. 2(d)) changes the tone noticeably. Contrast is stronger, the background is darker and more neutral, and fine structure, such as the beard, the folds of the wetsuit and the chest insignia, is crisper. A little grain shows up in the dark regions, which is the usual price of amplifying detail.

Fused output. The final image (Fig. 2(e) and Fig. 3) keeps the cleaner tone of Step 2 and adds the detail of Step 3. Facial features and the star insignia are clearly resolved, the diver's equipment shows structure, and the water is a believable teal-green and no longer a saturated green.

Table V condenses these observations. The entries are visual descriptions of Fig. 2 and are not additional measurements.

TABLE V
SUMMARY OF STAGE-WISE VISUAL EFFECTS

Stage	Observed visual effect (Fig. 2)
Input	Uniform bright green veil; subjects barely separable from water
Step 1	Green cast removed; bronze statue, blue mask frame, pale teal water
Step 2	Haze floor lowered; background slightly darker; mild gain in depth
Step 3	Stronger contrast and detail (beard, wetsuit folds, insignia); slight grain
Fused	Tone of Step 2 with detail of Step 3; facial and badge detail resolved

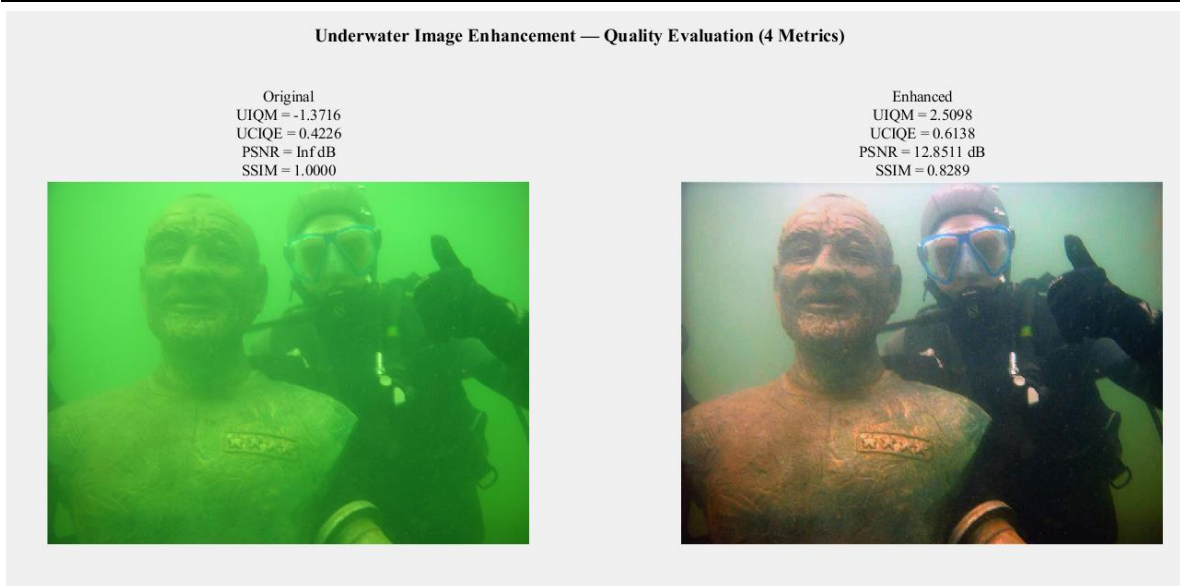


Fig. 3. Side-by-side comparison of the original (left) and the enhanced image (right), with metrics as reported by the pipeline. PSNR and SSIM are computed against the original image.

B. Weight-Map Analysis

Fig. 4 shows the three weight maps for each input. The Laplacian contrast weight is close to zero almost everywhere and reacts only to isolated edges; its colour scale reaches about 0.5–0.6. The saliency weight is smooth, with values up to roughly 4000 for input 1 and 2500 for input 2, strongest on the faces and the equipment. The saturation weight varies smoothly between 0 and about 0.5 and is highest in the lower part of the frame. The three ranges differ enormously.

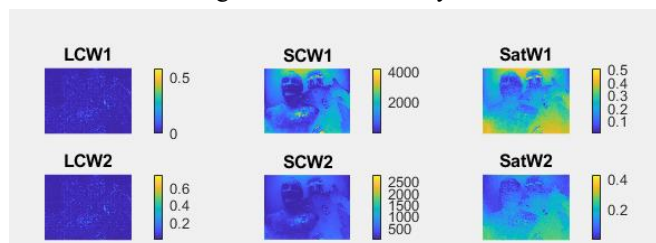


Fig. 4. Weight maps of Step 4 for the two fusion inputs (top row: input 1; bottom row: input 2). Left to right: Laplacian contrast (LCW), saliency (SCW), and saturation (SatW).

Table II confirms this. The mean saliency (1055.27) is about 4400 times the mean saturation (0.2404) and about 73 000 times the mean contrast (0.0145). Summed without normalisation, as in this run, the combined weight is effectively the saliency map alone, and the contrast and saturation terms hardly matter. We think this is a useful finding about the original weighting. It explains why the normalised weights in Fig. 5 are smooth with only weak structure from edges, and it is why the code offers an optional joint min–max normalisation that equalises the ranges before summation. We did not test that option here, so we cannot say it improves the result; the batch ablation mode is meant for exactly that test.

TABLE II
WEIGHT STATISTICS (NORMALISATION DISABLED)

Quantity	Mean value	Remark
LCW, input 1	0.0145	Laplacian contrast
SCW, input 1	1055.2727	Saliency (dominant)

Quantity	Mean value	Remark
SatW, input 1	0.2404	Saturation
NW ₁	0.7868	Input 1 (Step 2) share
NW ₂	0.2132	Input 2 (Step 3) share

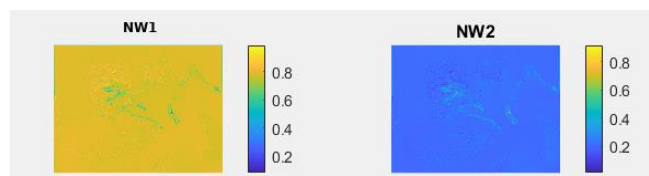


Fig. 5. Normalised fusion weights: NW1 (left, Step 2 input) and NW2 (right, Step 3 input). $NW_1 + NW_2 = 1$ at every pixel.

The mean normalised weights are $NW_1 = 0.7868$ and $NW_2 = 0.2132$ (Fig. 5). On average, input 1 (the scattering-compensated image) supplies about 79% of the fused low-frequency content and input 2 (the sharpened image) about 21%. NW1 is nearly uniform with local dips along edges and on the faces, where input 2 gains weight. That is the intended behaviour: the sharpened input contributes where local structure is rich. Input 1 may get the larger weight because the averaging in Step 3 narrows the spread of Lab values and so lowers its saliency, but we did not isolate this effect.

C. Quantitative Evaluation

Table III lists the metrics for the input and the enhanced image, and Fig. 6 plots them. The no-reference metrics improve clearly. UIQM rises from -1.3716 to 2.5098 , a gain of 3.8814 , and UCIQE rises from 0.4226 to 0.6138 , a gain of 0.1912 or 45.2% . Both agree with what the eye sees: the colour cast is gone and contrast is higher.

TABLE III
QUALITY METRICS FOR THE TEST IMAGE

Metric	Input	Enhanced	Delta	Type
UIQM	-1.3716	2.5098	$+3.8814$	No-ref.
UCIQE	0.4226	0.6138	$+0.1912$	No-ref.
PSNR (dB)	Inf*	12.8511	$-Inf^*$	Full-ref.*
SSIM	1.0000^*	0.8289	-0.1711^*	Full-ref.*

*No reference image was supplied; the input itself served as the reference, so these values measure change from the input.

Listing 1 reproduces the final summary exactly as printed by the program for this run. The verdict column is a simple sign test of the difference and does not account for the reference issue discussed in Section VII-D.

Listing 1. Final quality summary printed by the program

Metric	Input	Enhanced	Delta	Verdict
UIQM	-1.3716	2.5098	+3.8814	PASS
UCIQE	0.4226	0.6138	+0.1912	PASS
PSNR (dB)	Inf	12.8511	-Inf	CHECK
SSIM	1.0000	0.8289	-0.1711	CHECK

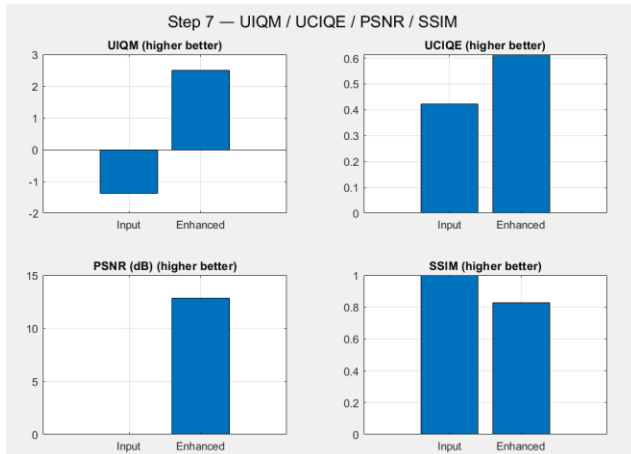


Fig. 6. Metric comparison for the input and the enhanced image. The input PSNR is infinite (identical to its own reference) and is therefore not drawn.

Since our UIQM variant is dominated by the $R - G$ term, a large part of the +3.88 gain simply reflects the correction of the green cast, the main defect of this image. UCIQE is less sensitive to hue and more to luminance contrast and saturation, so it gives independent support. The two metrics together are more convincing than either alone, but neither is ground truth.

Table IV summarises what each metric can and cannot establish in the present experiment. The distinction matters because the four numbers in Table III answer different questions: two describe the perceptual properties of an image taken in isolation, and two describe its distance from another image.

TABLE IV
INTERPRETATION OF THE FOUR METRICS IN THIS STUDY

Metric	Needs reference?	What it establishes here
UIQM	No	Colourfulness, sharpness, contrast of one image; dominated by $R - G$ term in this variant
UCIQE	No	Chroma variation, luminance contrast, saturation; independent check on UIQM
PSNR	Yes	Pixel-wise distance to the reference; here the distance from the input
SSIM	Yes	Structural similarity of greyscale images; here the structure kept from the input

D. Interpretation of PSNR and SSIM

PSNR and SSIM need careful reading in this run. No reference was selected, so the input served as its own reference. Its PSNR is therefore infinite and its SSIM exactly 1, and the

“delta” of $-\text{Inf}$ and the “CHECK” verdicts printed by the program come from this choice. They do not indicate failure. For the enhanced image, PSNR = 12.85 dB corresponds to an MSE of about 3370, a root-mean-square difference of about 58 grey levels (23% of full scale) from the input. SSIM = 0.8289 says the luminance structure is largely kept, with 17% structural change.

Put together, the numbers say the enhancement changed the intensities and colours a lot, as it should for a badly degraded image, while the geometry stayed intact. A low PSNR against a degraded input is in fact what any successful colour correction produces, and it should not be read as poor quality. The numbers do not tell how close the output is to the true scene. That needs a reference such as those in UIEB, and the batch mode of our program is designed to compute it.

E. Strengths

The pipeline has a few practical strengths. It needs no depth map, no calibration and no training data. Every stage can be displayed, which helps with debugging and teaching. The colour stage follows the dominant channel, so a strongly green image needs no special handling. Fusion keeps the detail, and the pyramid prevents visible seams. The metric module and batch mode also make it easy to extend the study to a full dataset.

In practice the pipeline suits situations where nothing is known about the camera or the water. A diver’s snapshot, a frame from a remotely operated vehicle or an archive photograph can be processed with the same settings. Because the intermediate images are available, a user who dislikes the final tone can pick the output of Step 2 or Step 3 instead, or change the fusion weights. We have not tested the pipeline on video, where frame-to-frame consistency of the colour statistics would become an issue, since Step 1 recomputes channel means for every frame.

F. Practical Considerations for the Main Parameters

We did not run a parameter sweep, but the structure of the pipeline allows some expectations, which readers can check with the provided scripts. The 0.5% clipping fraction trades robustness to outliers against the strength of the stretch: a larger fraction gives more contrast but clips more highlights and shadows, and the ratio-dependent scaling makes the effect strongest in the weakest channel. The exponent γ sets how far the haze floor is lowered. Values near 1 leave the image almost untouched, while large values darken mid-tones and may crush detail in dark regions such as the diving equipment. More pyramid levels give smoother tonal transitions but shrink the coarsest band and add a little cost, and δ controls how quickly the blend falls back to equal weights when raw weights are small. These are expectations drawn from the equations, not measured results.

G. Recommended Evaluation Protocol

From these observations we suggest the following reporting protocol. (1) Report the no-reference metrics for both the raw and the enhanced image, so the direction and size of the change are visible. (2) Report PSNR and SSIM only against a reference that was not derived from the input, and name the reference set. (3) Report mean and standard deviation over a dataset instead

of a single image, and state the weighting switch. (4) State which UIQM and UCIQE implementation was used, because several variants of the published formulas are in circulation. (5) Show the intermediate stages, since a metric gain can come mostly from one stage, as it does here with colour compensation.

H. Limitations

Several limitations should be stated plainly. (i) Only one image was used, so nothing can be said about average performance, robustness or significance. (ii) No competing method was run, so we make no claim of superiority. (iii) PSNR and SSIM were computed against the input and do not measure fidelity. (iv) Our UIQM and UCIQE are variants of the published formulas, so absolute values are not comparable with other papers. (v) With the unnormalised weighting, saliency dominates and the contrast and saturation weights contribute little. (vi) The final image has a warm orange-brown tone on the statue and a leftover teal-green veil in the upper background. The colour stage probably over-corrects the red channel in the foreground and under-corrects the distant water, which is typical for a global per-channel stretch that cannot vary with range. (vii) The image pyramid is built by decimation without pre-filtering, which can alias fine detail, and the gamma exponent and number of levels were fixed rather than adapted to the scene.

VIII. CONCLUSION AND FUTURE WORK

We presented a seven-step underwater enhancement pipeline that combines ratio-adaptive colour compensation, power-law scattering compensation, unsharp-mask contrast enhancement, and a three-level Laplacian-pyramid fusion with contrast, saliency and saturation weights, plus a built-in four-metric quality module. On a severely green-cast test image it removes the cast, brings back visible structure, and improves UIQM from -1.3716 to 2.5098 and UCIQE from 0.4226 to 0.6138 . The PSNR of 12.85 dB and SSIM of 0.8289 describe how much the image changed relative to the input, not fidelity to a reference, as we have said throughout.

There are four directions for future work. We will run the batch mode on a reference dataset such as UIEB for true full-reference statistics, reporting both weighting settings as an ablation. We will compare the pipeline with fusion-based, physics-based and learning-based baselines on common test sets, using the standard UIQM and UCIQE formulas alongside our variants. We want to make the colour stage range-aware, for example with a coarse transmission estimate that limits over-correction of the foreground. Finally, we will look at adaptive choices of the gamma exponent and pyramid depth, and at a pre-filtered pyramid, to make the method more robust across water types.

ACKNOWLEDGMENT

The authors thank [funding agency / institution] for support. The test image was used for research purposes only; please insert the image source and licence information before submission.

REFERENCES

- [1] J. S. Jaffe, "Computer modeling and the design of optimal underwater imaging systems," *IEEE J. Ocean. Eng.*, vol. 15, no. 2, pp. 101–111, Apr. 1990.
- [2] D. Akkaynak, T. Treibitz, T. Shlesinger, Y. Loya, R. Tamir, and D. Iluz, "What is the space of attenuation coefficients in underwater computer vision?" in *Proc. IEEE CVPR*, 2017, pp. 4931–4940.
- [3] D. Akkaynak and T. Treibitz, "A revised underwater image formation model," in *Proc. IEEE/CVF CVPR*, 2018, pp. 6723–6732.
- [4] J. Y. Chiang and Y.-C. Chen, "Underwater image enhancement by wavelength compensation and dehazing," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 1756–1769, Apr. 2012.
- [5] P. Drews Jr., E. do Nascimento, F. Moraes, S. Botelho, and M. Campos, "Transmission estimation in underwater single images," in *Proc. IEEE ICCV Workshops*, 2013, pp. 825–830.
- [6] 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.
- [7] K. Iqbal, R. A. Salam, A. Osman, and A. Z. Talib, "Underwater image enhancement using an integrated colour model," *IAENG Int. J. Comput. Sci.*, vol. 34, no. 2, pp. 239–244, 2007.
- [8] X. Fu, P. Zhuang, Y. Huang, Y. Liao, X.-P. Zhang, and X. Ding, "A retinex-based enhancing approach for single underwater image," in *Proc. IEEE ICIP*, 2014, pp. 4572–4576.
- [9] C. Li, S. Anwar, and F. Porikli, "Underwater scene prior inspired deep underwater image and video enhancement," *Pattern Recognit.*, vol. 98, Art. no. 107038, 2020.
- [10] C. Li *et al.*, "An underwater image enhancement benchmark dataset and beyond," *IEEE Trans. Image Process.*, vol. 29, pp. 4376–4389, 2020.
- [11] C. Ancuti, C. O. Ancuti, T. Haber, and P. Bekaert, "Enhancing underwater images and videos by fusion," in *Proc. IEEE CVPR*, 2012, pp. 81–88.
- [12] C. O. Ancuti, C. Ancuti, C. De Vleeschouwer, and P. Bekaert, "Color balance and fusion for underwater image enhancement," *IEEE Trans. Image Process.*, vol. 27, no. 1, pp. 379–393, Jan. 2018.
- [13] K. He, J. Sun, and X. Tang, "Single image haze removal using dark channel prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 12, pp. 2341–2353, Dec. 2011.
- [14] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press, 1994, pp. 474–485.
- [15] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, no. 4, pp. 532–540, Apr. 1983.
- [16] T. Mertens, J. Kautz, and F. Van Reeth, "Exposure fusion: A simple and practical alternative to high dynamic range photography," *Comput. Graph. Forum*, vol. 28, no. 1, pp. 161–171, 2009.
- [17] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunk, "Frequency-tuned salient region detection," in *Proc. IEEE CVPR*, 2009, pp. 1597–1604.
- [18] K. Panetta, C. Gao, and S. Agaian, "Human-visual-system-inspired underwater image quality measures," *IEEE J. Ocean. Eng.*, vol. 41, no. 3, pp. 541–551, Jul. 2016.
- [19] M. Yang and A. Sowmya, "An underwater color image quality evaluation metric," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 6062–6071, Dec. 2015.
- [20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.