

# Colorization for *in situ* marine plankton images

## Supplementary Material

Guannan Guo<sup>1,2</sup>, Qi Lin<sup>3</sup>, Tao Chen<sup>1,2</sup>, Zhenghui Feng<sup>4</sup>, Zheng Wang<sup>1,2</sup>, and Jianping Li<sup>1,2</sup>✉

<sup>1</sup> Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

[jp.li@siat.ac.cn](mailto:jp.li@siat.ac.cn)

<sup>2</sup> University of Chinese Academy of Sciences, Beijing, China

<sup>3</sup> Xiamen University, Xiamen, China

<sup>4</sup> Harbin Institute of Technology, Shenzhen, China

## 1 Details for Model Training and Ablation Experiments

In the dataset for IsPlanktonCLR model training, the images in the Training Set and Testing Set 1 are color ROIs. We use their  $L$  channel as the input for training network and  $ab$  channel as the ground truth to validate the network, respectively. The Testing Set 2 contains 60 grayscale-color image pairs. The grayscale images are used as network input to generate colorized images, and their color counterparts are used as ground truth to validate the colorization performance.

We train the IsPlanktonCLR network for 600 epochs with batch size of 16 on an NVIDIA Tesla A100 GPU. We use Adam optimizer with an initial learning rate 0.001, which decays every 400 epochs. The input image size is standardized to  $224 \times 224$  pixels as done in [8].

**Table 1.** Ablation experiments and numerical comparison with other SOTA methods.

	backbone	Net1( $\omega_0, 1$ )	Net2( $\omega_1, 1$ )	Net3( $\omega_1, 2$ )	Ours(whole)	InstCol	Chroma
CDSIM↓	463.917	457.059	455.631	454.420	<b>346.434</b>	651.904	442.487
FID↓	35.754	32.285	30.768	31.420	<b>24.578</b>	45.397	47.123
PSNR↑	42.491	43.414	42.660	40.481	<b>44.269</b>	42.535	40.431
SSIM↑	0.993	0.994	0.991	0.987	<b>0.996</b>	0.984	0.970

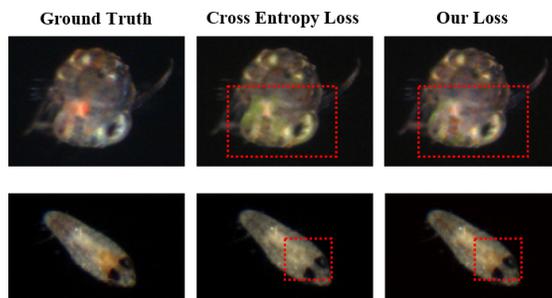
We perform a set of ablation experiments to show the impact of each module in IsplanktonCLR. Table 1 summarizes the effect of each module. The **backbone** only has **Extractor** and **Decoder**, **Net1~3** adds our palette and loss function, and **Ours** adds **Classifier** and **Encoder**. It can be seen that with the self-guidance reference, the metrics of CDSIM and FID that can reflect the colourisation accuracy have been significantly optimized. The simplification of the palette and the improvement of the loss function can reduce the complexity of the model and accelerate the convergence speed.

In addition, we also compare two SOTA models InstCol[13] and Chroma[15] in Table 1. The results are consistent with our analysis.

## 2 Details for Loss Function

The value determination of  $\omega$  and  $\gamma$  in Eq.4 of our loss is as follows. For  $\omega$ , we assign equal weight  $\omega_0$  to each colour, and weight  $\omega_1$  to each colour according to its proportion in the training set. The larger the proportion, the smaller the weight. Specifically, we separate the quantity proportion into four intervals of  $[1, 0.1]$ ,  $(0.1, 0.01]$ ,  $(0.01, 0.001]$ , and  $(0.001, 0]$ , corresponding to weights of 0.001, 0.009, 0.09, and 0.9, respectively. Assuming that there are  $N$  colours in  $[1, 0.1]$ , the weight of each colour is  $0.001/N$ . For  $\gamma$ , we test the case with values of 1 and 2, and the results do not change significantly.

Fig. 1 compares the colorization effect between two IsPlanktonCLR models trained with the baseline cross entropy loss and our proposed loss after 100 epochs, respectively. We find that our loss can make the model converge faster under the same conditions than the cross entropy loss does. As more detailed observation indicated in the red dotted boxes, the model trained with our loss is able to correctly recover more reddish color for the key parts of the plankton targets, while the model using the cross entropy loss presents quite some wrong greenish colors in the top example or is unable to recover reddish colors in the bottom example.



**Fig. 1.** Colorization effect comparison between IsPlanktonCLR models trained with different losses.

## 3 Extensive Representation of *in situ* Plankton Image Colorization

To demonstrate the colorization performance of IsPlanktonCLR on more diverse plankton taxa, we select 84 new examples from different classes of the

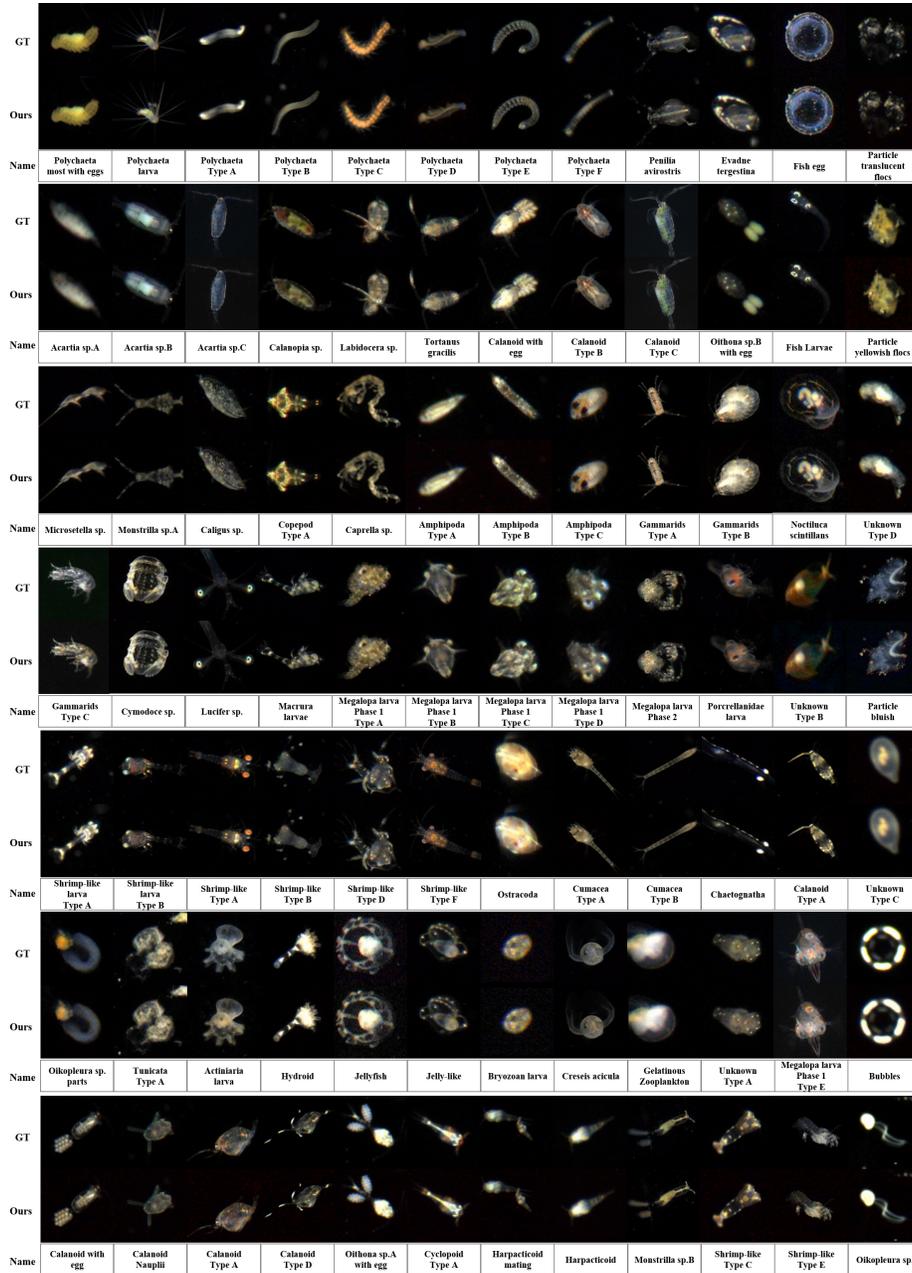


Fig. 2. 84 classes of representative marine plankton and suspended particles ROI images from the DYBPlanktonNet [9] colorized by the IsPlanktonCLR model.

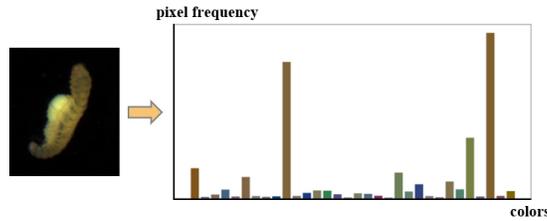
DYB-PlanktonNet [9] dataset and extract their  $L$  channel as grayscale image input fed into the model. Their colorization results are gathered in Fig. 2. The results show that IsPlanktonCLR can restore most examples to their original colors perceptually, which indicates that the algorithm can achieve consistent colorization performance at least on plankton taxa in specific sea areas such as Daya Bay.

## 4 Calculation of Color Dissimilarity Metric

We first introduce the definitions and calculation of color histogram (CH) [5,11,12], color coherence vector (CCV) [10], color correlogram (CC) [6] and color gradient (CG) [3] for color feature extraction, and then introduce the regularization and dimensionality reduction methods for them to calculate the metric of CDSIM.

### 4.1 Color Feature Definition and Calculation

**Color Histogram (CH)** is a statistical histogram formed by counting the frequency of each color occurrence in a color image. CH reflects the proportion of different colors in an image, but cannot represent the spatial distribution of colors. Fig. 3 shows the CH of a plankton image example.



**Fig. 3.** Color histogram of a ROI image containing a Polychaete.

**Color Coherence Vector (CCV)** is a color feature containing the spatial distribution information of colors. If a continuous area in a color image occupied by some pixels of the same color is larger than a given threshold, the pixels in this region are called coherent pixels; otherwise, they are regarded as incoherent pixels [10]. CCV adds up the number of coherent and incoherent pixels of the same color in a CH, and divides pixels of each color into coherent and incoherent parts. Suppose the color image contains a total of  $N$  colors,  $\alpha_i$  and  $\beta_i$  represent respectively the number of coherent pixels and incoherent pixels of the  $i^{th}$  color in its CH, then  $(\alpha_i, \beta_i)$  is called a Coherence Pair, and the CCV of the image is expressed as:

$$\langle (\alpha_1, \beta_1), (\alpha_2, \beta_2), \dots, (\alpha_N, \beta_N) \rangle . \quad (1)$$

**Color Correlogram (CC)** is another color feature incorporates spatial distribution of colors. It considers the distribution of various colors around a color. Suppose there are  $N$  colors in an image, denoted as  $c_1, \dots, c_N$ ; the coordinates of a pixel of the image are denoted as  $I = (x, y)$  with its color  $C(I)$ ; the coordinates of all pixels on the image is a set of  $J$ , and all the pixel coordinates corresponding to color  $c_i$  is denoted as a set  $J_{c_i} \triangleq \{I | C(I) = c_i\}$ . Then, we can define the probability of color  $c_j$  at a distance  $d$  to  $c_i$  as:

$$\gamma_{c_i, c_j}^{(d)}(J) \triangleq P_r[I_2 \in J_{c_j} | \text{dist}(I_1, I_2) = d, I_1 \in J_{c_i}, I_2 \in J], i = 1, 2, \dots, N, \quad (2)$$

where  $\text{dist}(I_1, I_2)$  is the distance between pixels of  $p_1$  and  $p_2$  measured by the  $L_\infty$  norm as :

$$\text{dist}(I_1, I_2) \triangleq \max\{|x_1 - x_2|, |y_1 - y_2|\}, \quad (3)$$

and  $d \leq d_0$ ,  $d_0$  is a fixed *a priori*.  $\gamma_{c_i, c_j}^{(d)}(J)$  represents the color distribution along the four edges of a square with its center of color  $c_i$  and side length of  $d$ . If  $d = 1$ ,  $\gamma_{c_i, c_j}^{(1)}(J)$  represents the probability of color  $c_j$  in the eight surrounding pixels for all pixels with color  $c_i$ .

We use the pixel frequency to estimate  $\gamma_{c_i, c_j}^{(d)}(J)$ , and the CC at distance  $d$  can be written as a square matrix of order  $N$ :

$$\hat{\gamma}^{(d)} = \begin{pmatrix} \hat{\gamma}_{c_1, c_1}^{(d)}(J) & \hat{\gamma}_{c_1, c_2}^{(d)}(J) & \dots & \hat{\gamma}_{c_1, c_N}^{(d)}(J) \\ \hat{\gamma}_{c_2, c_1}^{(d)}(J) & \hat{\gamma}_{c_2, c_2}^{(d)}(J) & \dots & \hat{\gamma}_{c_2, c_N}^{(d)}(J) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}_{c_N, c_1}^{(d)}(J) & \hat{\gamma}_{c_N, c_2}^{(d)}(J) & \dots & \hat{\gamma}_{c_N, c_N}^{(d)}(J) \end{pmatrix}, \quad (4)$$

where  $d = 1, 2, \dots, d_0$ . It can be seen that CC has a total of  $N^2 d_0$  elements.

The dimensionality of CC is much greater than that of CCV and CH ( $N^2 d_0 \geq 2N \geq N$ ). We only use the elements on the diagonal of  $\hat{\gamma}^{(d)}$  as CC feature, which reduces the dimensionality from  $N^2 d_0$  to  $N d_0$ . The diagonal elements are also called Color Auto-correlogram (CAC), which can be expressed as:

$$\hat{\gamma}_{c_i, c_i}^{(d)}(J), i = 1, 2, 3, \dots, N. \quad (5)$$

**Color Gradient (CG)** stands for gradient features calculated using color information from a multi-channel image. We use two methods to compute CG features from RGB color images, respectively.

The first method uses *Sobel* operator [4] to obtain gradients from a single color channel. First, we calculate the gradients  $R_x$ ,  $R_y$ ,  $G_x$ ,  $G_y$ ,  $B_x$ , and  $B_y$  along the  $x$  and  $y$  directions at the pixel position of  $(x, y)$  on each of the  $R$ ,  $G$ ,  $B$  channels of a color image, and then calculate the CG feature as:

$$M(x, y) = \sqrt{\frac{R_x^2 + R_y^2}{2} + \frac{G_x^2 + G_y^2}{2} + \frac{B_x^2 + B_y^2}{2}}. \quad (6)$$

This method considers the gradients of  $R$ ,  $G$ , and  $B$  channels along the  $x$  and  $y$  directions, but ignores the gradient information between color channels.

The second method considers the relationship between different channels [3]. It defines the direction along which the gradient is maximum and is expressed as:

$$\theta(x, y) = \frac{1}{2} \tan^{-1} \left[ \frac{2g_{xy}}{g_{xx} - g_{yy}} \right], \quad (7)$$

where

$$g_{xx} = \mathbf{u} \cdot \mathbf{u} = \mathbf{u}^T \mathbf{u} = \left| \frac{\partial R}{\partial x} \right|^2 + \left| \frac{\partial G}{\partial x} \right|^2 + \left| \frac{\partial B}{\partial x} \right|^2, \quad (8)$$

$$g_{yy} = \mathbf{v} \cdot \mathbf{v} = \mathbf{v}^T \mathbf{v} = \left| \frac{\partial R}{\partial y} \right|^2 + \left| \frac{\partial G}{\partial y} \right|^2 + \left| \frac{\partial B}{\partial y} \right|^2, \quad (9)$$

$$g_{xy} = \mathbf{u} \cdot \mathbf{v} = \mathbf{u}^T \mathbf{v} = \frac{\partial R}{\partial x} \frac{\partial R}{\partial y} + \frac{\partial G}{\partial x} \frac{\partial G}{\partial y} + \frac{\partial B}{\partial x} \frac{\partial B}{\partial y}, \quad (10)$$

and

$$\mathbf{u} = \frac{\partial R}{\partial x} \mathbf{r} + \frac{\partial G}{\partial x} \mathbf{g} + \frac{\partial B}{\partial x} \mathbf{b}, \quad (11)$$

$$\mathbf{v} = \frac{\partial R}{\partial y} \mathbf{r} + \frac{\partial G}{\partial y} \mathbf{g} + \frac{\partial B}{\partial y} \mathbf{b}. \quad (12)$$

Here  $\mathbf{r}$ ,  $\mathbf{g}$ , and  $\mathbf{b}$  are unitary vectors associated with the  $R$ ,  $G$ , and  $B$  axes, respectively.

Thus, the CG feature can be calculated as:

$$F_\theta(x, y) = \sqrt{\frac{1}{2} [(g_{xx} + g_{yy}) + (g_{xx} - g_{yy}) \cos 2\theta(x, y) + 2g_{xy} \sin 2\theta(x, y)]}. \quad (13)$$

## 4.2 Color Feature Dimension Reduction

The original color features introduced above are not regularized and with very high dimensions. For ease of CDISM computation, we use the following two approaches to regularize them and reduce their dimensionality.

The first approach is to use statistical descriptors of the single-channel histogram (SCH) for dimensionality reduction. Specifically, we first get the SCHs of a color image, and then the statistical descriptors listed in Table 2 are calculated based on the formulas listed on the right [1]. Among them, CH, CCV, and CAC features are calculated in three color spaces of RGB, LCH, and HSV, respectively; and CG features are calculated in the RGB color space.

The second approach is to extract CH, CCV, and CAC features from a simplified color palette of the color image. Specifically, the  $K$ -means++ algorithm

**Table 2.** Statistical descriptors extracted from a single-channel histogram.

Names	Formulas
Mean	$\mu = \sum_{i=0}^{H-1} p(i) \cdot h(i)$
Mode	$i = \operatorname{argmax}(h(i))$
Minimum	$\min(h(i))$
Maximum	$\max(h(i))$
Variance	$\sigma^2 = \sum_{i=0}^{H-1} (p(i) - \mu)^2 \cdot h(i)$
Range	$\max(h(i)) - \min(h(i))$
Entropy	$-\sum_{i=0}^{H-1} h(i) \cdot \log(h(i))$
1st Quartile Mean	$\mu_{q1} = \sum_{i=\lceil 3H/4 \rceil}^{H-1} p(i) \cdot h(i)$
2nd Quartile Mean	$\mu_{q2} = \sum_{i=\lceil H/2 \rceil}^{\lceil 3H/4 \rceil} p(i) \cdot h(i)$
3rd Quartile Mean	$\mu_{q3} = \sum_{i=\lceil H/4 \rceil}^{\lceil H/2 \rceil} p(i) \cdot h(i)$
Quartile Mean Difference	$\mu_{q1} - \mu_{q3}$
Upper Quartile	$I_{(M/4)}$
Lower Quartile	$I_{(3M/4)}$
Quartile Difference	$I_{(3M/4)} - I_{(M/4)}$
Median	$I_{(M/2)}$
Asymmetry(Skewness)	$\frac{1}{\sigma^3} \sum_{i=0}^{H-1} (p(i) - \mu)^3 \cdot h(i)$
Kurtosis	$\frac{1}{\sigma^4} \sum_{i=0}^{H-1} (p(i) - \mu)^4 \cdot h(i)$

$H$  is the number of bins in the histogram,  $p(i)$  is the pixel value of the  $i^{th}$  bin,  $h(i)$  is the value frequency of the  $i^{th}$  bin,  $I_{(j)}$  is the  $j^{th}$  largest pixel value in an image, and  $M$  is the pixel number.

[14,2] described in the Section 3.1 of the manuscript is firstly used to cluster the image colors on the HSV, Lab, and Luv color spaces, respectively. Then the CH, CCV, and CAC features are extracted from the simplified color palette of the original image.

We summarize the number of all the color features used for calculating CD-SIM metric in this work in Table 3.

**Table 3.** Summary of color features used for CDSIM computation in this work.

Features	SCH	CH	CCV	CAC	CG	Total
Number	153	96	192	768	51	1260

## 5 Online Survey for Visual Evaluation of Colorization Effect

We present the detailed design and results of the questionnaire used in this work. The questionnaire consists of 14 questions. The participants are asked to evaluate the color similarity between the targets in the ROI.1 ~ ROI.4 and the GT in each question. ROI.1 ~ ROI.4 in each question corresponds to the colorization results of the  $L$  channel grayscale image of the GT by IsPlanktonCLR, MemoColor [16], LetColor [7], and CIC [17] models, respectively. The scores of color similarity are scaled in 5 points with 5 being the most similar and 1 being the least similar.

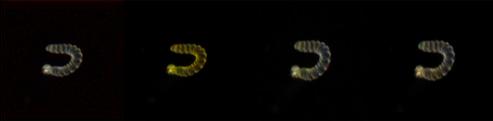
The questionnaire was sent to three WeChat groups consisting of 45 PhD students, 490 marine biologists, and 7 plankton taxonomists. 115 valid results from participants without any knowledge of the background and intension of this survey were received within two days before the submission of this manuscript.

The table below each question details the vote and score results for that set of image comparison. Finally, the average scores for the 14 questions are calculated and listed in Table 2 in the main manuscript. The results of this survey prove that the colorization effect of IsPlanktonSR achieves the closest to extensive human visual perception evaluation among all the SOTA colorization models.

## Question 1.

GT	ROI1	ROI2	ROI3	ROI4		
						
ROI \ Similarity	1	2	3	4	5	Average
ROI.1	10(8.7%)	58(50.43%)	31(26.96%)	7(6.09%)	9(7.83%)	2.54
ROI.2	66(57.39%)	25(21.74%)	9(7.83%)	9(7.83%)	6(5.22%)	1.82
ROI.3	1(0.87%)	20(17.39%)	45(39.13%)	42(36.52%)	7(6.09%)	3.3
ROI.4	3(2.61%)	4(3.48%)	12(10.43%)	46(40%)	50(43.48%)	<b>4.18</b>

## Question 2.

GT	ROI1	ROI2	ROI3	ROI4		
						
ROI \ Similarity	1	2	3	4	5	Average
ROI.1	2(1.74%)	8(6.96%)	15(13.04%)	34(29.57%)	56(48.7%)	<b>4.17</b>
ROI.2	69(60%)	16(13.91%)	13(11.3%)	11(9.57%)	6(5.22%)	1.86
ROI.3	6(5.22%)	25(21.74%)	44(38.26%)	34(29.57%)	6(5.22%)	3.08
ROI.4	9(7.83%)	39(33.91%)	41(35.65%)	22(19.13%)	4(3.48%)	2.77

## Question 3.

GT	ROI1	ROI2	ROI3	ROI4		
						
ROI \ Similarity	1	2	3	4	5	Average
ROI.1	22(19.13%)	28(24.35%)	38(33.04%)	16(13.91%)	11(9.57%)	2.7
ROI.2	54(46.96%)	22(19.13%)	16(13.91%)	14(12.17%)	9(7.83%)	2.15
ROI.3	11(9.57%)	25(21.74%)	44(38.26%)	26(22.61%)	9(7.83%)	2.97
ROI.4	5(4.35%)	23(20%)	34(29.57%)	32(27.83%)	21(18.26%)	<b>3.36</b>

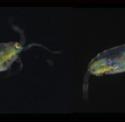
**Question 4.**

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI\ Similarity	1	2	3	4	5	Average
ROI1	35(30.43%)	38(33.04%)	23(20%)	12(10.43%)	7(6.09%)	2.29
ROI2	34(29.57%)	36(31.3%)	22(19.13%)	15(13.04%)	8(6.96%)	2.37
ROI3	4(3.48%)	15(13.04%)	46(40%)	34(29.57%)	16(13.91%)	<b>3.37</b>
ROI4	9(7.83%)	22(19.13%)	41(35.65%)	32(27.83%)	11(9.57%)	3.12

**Question 5.**

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI\ Similarity	1	2	3	4	5	Average
ROI1	1(0.87%)	4(3.48%)	17(14.78%)	45(39.13%)	48(41.74%)	<b>4.17</b>
ROI2	47(40.87%)	25(21.74%)	20(17.39%)	17(14.78%)	6(5.22%)	2.22
ROI3	48(41.74%)	38(33.04%)	19(16.52%)	8(6.96%)	2(1.74%)	1.94
ROI4	56(48.7%)	32(27.83%)	22(19.13%)	4(3.48%)	1(0.87%)	1.8

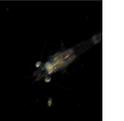
**Question 6.**

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI\ Similarity	1	2	3	4	5	Average
ROI1	2(1.74%)	4(3.48%)	7(6.09%)	33(28.7%)	69(60%)	<b>4.42</b>
ROI2	5(4.35%)	6(5.22%)	25(21.74%)	59(51.3%)	20(17.39%)	3.72
ROI3	7(6.09%)	23(20%)	58(50.43%)	20(17.39%)	7(6.09%)	2.97
ROI4	32(27.83%)	43(37.39%)	21(18.26%)	14(12.17%)	5(4.35%)	2.28

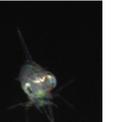
## Question 7.

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI\ Similarity	1	2	3	4	5	Average
ROI1	2(1.74%)	8(6.96%)	12(10.43%)	50(43.48%)	43(37.39%)	<b>4.08</b>
ROI2	60(52.17%)	22(19.13%)	17(14.78%)	12(10.43%)	4(3.48%)	1.94
ROI3	10(8.7%)	42(36.52%)	39(33.91%)	21(18.26%)	3(2.61%)	2.7
ROI4	9(7.83%)	36(31.3%)	35(30.43%)	31(26.96%)	4(3.48%)	2.87

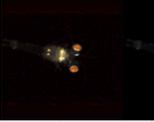
## Question 8.

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI\ Similarity	1	2	3	4	5	Average
ROI1	2(1.74%)	8(6.96%)	29(25.22%)	47(40.87%)	29(25.22%)	<b>3.81</b>
ROI2	4(3.48%)	16(13.91%)	45(39.13%)	39(33.91%)	11(9.57%)	3.32
ROI3	9(7.83%)	42(36.52%)	48(41.74%)	12(10.43%)	4(3.48%)	2.65
ROI4	48(41.74%)	43(37.39%)	15(13.04%)	3(2.61%)	6(5.22%)	1.92

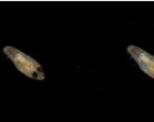
## Question 9.

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI\ Similarity	1	2	3	4	5	Average
ROI1	3(2.61%)	3(2.61%)	2(1.74%)	34(29.57%)	73(63.48%)	<b>4.49</b>
ROI2	35(30.43%)	35(30.43%)	31(26.96%)	10(8.7%)	4(3.48%)	2.24
ROI3	10(8.7%)	39(33.91%)	45(39.13%)	18(15.65%)	3(2.61%)	2.7
ROI4	27(23.48%)	45(39.13%)	28(24.35%)	11(9.57%)	4(3.48%)	2.3

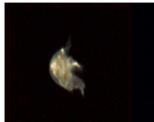
**Question 10.**

GT	ROI1	ROI2	ROI3	ROI4		
						
ROI\ Similarity	1	2	3	4	5	Average
ROI.1	3(2.61%)	2(1.74%)	4(3.48%)	23(20%)	83(72.17%)	<b>4.57</b>
ROI.2	4(3.48%)	8(6.96%)	37(32.17%)	60(52.17%)	6(5.22%)	3.49
ROI.3	10(8.7%)	28(24.35%)	55(47.83%)	20(17.39%)	2(1.74%)	2.79
ROI.4	12(10.43%)	38(33.04%)	51(44.35%)	11(9.57%)	3(2.61%)	2.61

**Question 11.**

GT	ROI1	ROI2	ROI3	ROI4		
						
ROI\ Similarity	1	2	3	4	5	Average
ROI.1	3(2.61%)	1(0.87%)	4(3.48%)	40(34.78%)	67(58.26%)	<b>4.45</b>
ROI.2	4(3.48%)	11(9.57%)	31(26.96%)	46(40%)	23(20%)	3.63
ROI.3	5(4.35%)	14(12.17%)	27(23.48%)	47(40.87%)	22(19.13%)	3.58
ROI.4	6(5.22%)	12(10.43%)	21(18.26%)	50(43.48%)	26(22.61%)	3.68

**Question 12.**

GT	ROI1	ROI2	ROI3	ROI4		
						
ROI\ Similarity	1	2	3	4	5	Average
ROI.1	8(6.96%)	25(21.74%)	38(33.04%)	28(24.35%)	16(13.91%)	3.17
ROI.2	33(28.7%)	32(27.83%)	30(26.09%)	12(10.43%)	8(6.96%)	2.39
ROI.3	5(4.35%)	23(20%)	30(26.09%)	38(33.04%)	19(16.52%)	<b>3.37</b>
ROI.4	15(13.04%)	32(27.83%)	36(31.3%)	25(21.74%)	7(6.09%)	2.8

## Question 13.

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI Similarity	1	2	3	4	5	Average
ROI1	4(3.48%)	15(13.04%)	14(12.17%)	44(38.26%)	38(33.04%)	<b>3.84</b>
ROI2	29(25.22%)	38(33.04%)	23(20%)	21(18.26%)	4(3.48%)	2.42
ROI3	22(19.13%)	36(31.3%)	35(30.43%)	15(13.04%)	7(6.09%)	2.56
ROI4	51(44.35%)	24(20.87%)	20(17.39%)	13(11.3%)	7(6.09%)	2.14

## Question 14.

	GT	ROI1	ROI2	ROI3	ROI4	
						
ROI Similarity	1	2	3	4	5	Average
ROI1	3(2.61%)	4(3.48%)	7(6.09%)	44(38.26%)	57(49.57%)	<b>4.29</b>
ROI2	2(1.74%)	5(4.35%)	33(28.7%)	51(44.35%)	24(20.87%)	3.78
ROI3	7(6.09%)	33(28.7%)	38(33.04%)	28(24.35%)	9(7.83%)	2.99
ROI4	38(33.04%)	39(33.91%)	22(19.13%)	10(8.7%)	6(5.22%)	2.19

## References

1. Bueno, G., Deniz, O., Pedraza, A., Ruiz-Santaquiteria, J., Salido, J., Cristóbal, G., Borrego-Ramos, M., Blanco, S.: Automated diatom classification (part a): hand-crafted feature approaches. *Applied Sciences* **7**(8), 753 (2017)
2. Celebi, M.E.: Improving the performance of k-means for color quantization. *Image and Vision Computing* **29**(4), 260–271 (2011)
3. Di Zenzo, S.: A note on the gradient of a multi-image. *Computer vision, graphics, and image processing* **33**(1), 116–125 (1986)
4. Gonzalez, R.C.: *Digital image processing*. Pearson education india (2009)
5. Hsu, W., Chua, S., Pung, H.: An integrated color-spatial approach to content-based image retrieval. In: *Proceedings of the third ACM international conference on Multimedia*. pp. 305–313 (1995)
6. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Image indexing using color correlograms. In: *Proceedings of IEEE computer society conference on Computer Vision and Pattern Recognition*. pp. 762–768. IEEE (1997)

7. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Let there be color! joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification. *ACM Transactions on Graphics (ToG)* **35**(4), 1–11 (2016)
8. Li, J., Chen, T., Yang, Z., Chen, L., Liu, P., Zhang, Y., Yu, G., Chen, J., Li, H., Sun, X.: Development of a buoy-borne underwater imaging system for in situ mesoplankton monitoring of coastal waters. *IEEE Journal of Oceanic Engineering* **47**(1), 88–110 (2021)
9. Li, J., Yang, Z., Chen, T.: Dyb-planktonnet. *IEEE Dataport* (2021)
10. Pass, G., Zabih, R., Miller, J.: Comparing images using color coherence vectors. In: *Proceedings of the fourth ACM international conference on Multimedia*. pp. 65–73 (1997)
11. Rickman, R.M., Stonham, T.J.: Content-based image retrieval using color tuple histograms. In: *Storage and Retrieval for Still Image and Video Databases IV*. vol. 2670, pp. 2–7. International Society for Optics and Photonics (1996)
12. Stricker, M.A., Dimai, A.: Color indexing with weak spatial constraints. In: *Storage and Retrieval for Still Image and Video Databases IV*. vol. 2670, pp. 29–40. SPIE (1996)
13. Su, J.W., Chu, H.K., Huang, J.B.: Instance-aware image colorization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 7968–7977 (2020)
14. Vailaya, A., Jain, A., Zhang, H.J.: On image classification: City images vs. landscapes. *Pattern recognition* **31**(12), 1921–1935 (1998)
15. Vitoria, P., Raad, L., Ballester, C.: Chromagan: Adversarial picture colorization with semantic class distribution. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. pp. 2445–2454 (2020)
16. Yoo, S., Bahng, H., Chung, S., Lee, J., Chang, J., Choo, J.: Coloring with limited data: Few-shot colorization via memory augmented networks. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 11283–11292 (2019)
17. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: *European conference on computer vision*. pp. 649–666. Springer (2016)