

# Colorization for *in situ* marine plankton images

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

<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

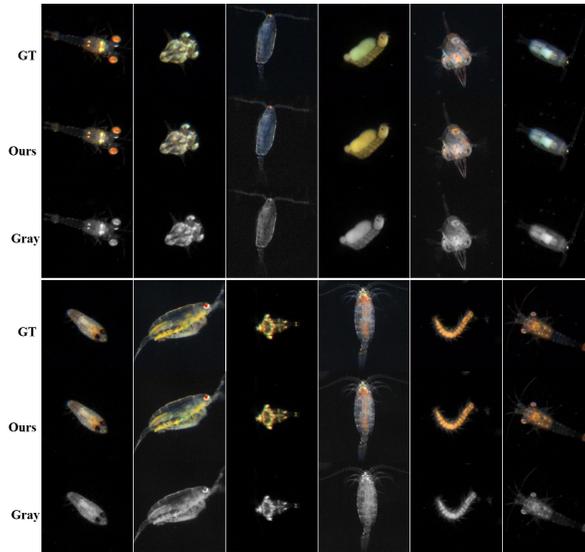
**Abstract.** Underwater imaging with red-NIR light illumination can avoid phototropic aggregation-induced observational deviation of marine plankton abundance under white light illumination, but this will lead to the loss of critical color information in the collected grayscale images, which is non-preferable to subsequent human and machine recognition. We present a novel deep networks-based vision system IsPlanktonCLR for automatic colorization of *in situ* marine plankton images. IsPlanktonCLR uses a reference module to generate self-guidance from a customized palette, which is obtained by clustering *in situ* plankton image colors. With this self-guidance, a parallel colorization module restores input grayscale images into their true color counterparts. Additionally, a new metric for image colorization evaluation is proposed, which can objectively reflect the color dissimilarity between comparative images. Experiments and comparisons with state-of-the-art approaches are presented to show that our method achieves a substantial improvement over previous methods on color restoration of scientific plankton image data.

**Keywords:** image colorization, deep learning, underwater imaging, *in situ* observation, marine plankton

## 1 Introduction

*In situ* imaging of marine plankton has been demonstrated very promising for scientific research to understand marine ecosystems and also become appealing for modern ocean management [25,36]. Limited by working principle and device performance, most early underwater plankton cameras can only capture grayscale images [5]. With recent technology development, some dark-field underwater cameras have been enabled for color imaging [13,6,21,33]. The color images captured by them have been shown to improve a machine classifier’s accuracy than that achieved on grayscale ones [21]. However, color imaging requires white light illumination, which causes phototropic aggregation of zooplankton frequently, thus resulting in plankton abundance measurement bias and great concern for observation accuracy [38,39].

Since most zooplankton are insensitive to longer wavelengths [12], *in situ* cameras can use red-NIR light for illumination to avoid phototropic aggregation [9,32]. But using red-NIR lighting will make an underwater camera acquire just grayscale images. If such grayscale images can be colorized with high-fidelity using deep learning techniques [1], it would not only enrich the image data with extra color information to facilitate subsequent human and machine recognition, but can also completely resolve the concern on plankton abundance measurement error associated with white light illumination. Moreover, a grayscale image not only has higher spatial resolution than its color counterpart with the same pixel size and number, its file size is also much smaller than the color version. This is beneficial to reduce the resource stress of image data processing, storage, and transmission for achieving sustainable ocean observation.



**Fig. 1.** Representative examples of *in situ* marine plankton grayscale images, their colorizations by IsPlanktonCLR, and the ground truth.

The demand for *in situ* plankton grayscale image colorization obviously corresponds to an image restoration problem. However, existing colorization algorithms are mainly developed for colorizing natural scene images captured in the real world [8,17,41,40,7,30,4], which corresponds essentially to an image enhancement task. As the colorization of these images is more in pursuit of rationality, comfort and diversity of human visual perception, the same target can be artificially painted into multiple colors. For example, a blue T-shirt can be colorized into a green, or a yellow or a red one. This task is an ill-posed problem, in which the deep colorization networks are difficult to establish deterministic mapping between the input and the output, and hence unable to meet higher demands

in colorization accuracy for scientific imaging applications. On the other hand, many restoration-oriented algorithms need user guidance to ensure the colorization effect [11,15]. However, these guidance is either given by human in advance or obtained through human-computer interaction, which is not conducive to their applications in long-term and automated ocean observation activities.

On this regard, we treat it as a color classification problem and propose a self-guided automatic deep colorization algorithm for color restoration of *in situ* plankton grayscale images. The network is named IsPlanktonCLR, whose idea and architecture is illustrated in Figure 2. In the design of this network, we firstly customize a reference palette to reduce the number of colors in the searching space, so as to achieve satisfying colorization effect with better efficiency. Then, to further ensure colorization accuracy and avoid color averaging effect, we combine the advantages of both user-guided and big-data driven algorithms to improve the model performance of color restoration in a self-guided and automatic way. This is achieved through a parallel network architecture consisting of a primary module for image colorization and an additional reference module for providing guidance from the customized palette. Using this method, we successfully achieve satisfying colorization effect on the *in situ* marine plankton image data as shown in Figure 1.

The feasibility of IsPlanktonCLR is based on the premise that the coloration of marine plankton is relatively monotonous in their *in situ* images. By investigating human visual perception of a large number of images, we realize that most plankton only show one or two families of colors. This allows us to group the plankton images based on their color families, and use this information to label them as references for guiding the colorization.

We are also aware that the obvious imbalance in plankton image dataset leads to the imbalance in the color quantities. For example, there are a lot of yellowish colors in the dataset, while the reddish colors are less common. Direct use of imbalanced data to train the model will cause dominant color effect, which is a known problem associated with many deep colorization models [44]. To resolve this issue, we use data augmentation and loss reweighting to enforce the model learning more on rare colors.

In addition, we notice an obvious lack of objective and quantitative colorization evaluation metrics for restoration-oriented scientific image colorization. Although PSNR, SSIM and other metrics are often used for image restoration evaluation, they are known to be less effective for colorization evaluation [17,40,47,7,29,37], and are often inconsistent with human perception. Therefore, we propose a new metric Color Dissimilarity (CDISM) to better characterize the color restoration accuracy of an output image relative to its ground truth. CDSIM is obtained by calculating the Euclidean distance between color feature vectors extracted from two comparative images. We demonstrate its effectiveness on both plankton and natural scene images.

To summarize, the contribution of this work includes:

—A new idea in automatic grayscale image colorization for scientific domain imagery is provided, which improves the colorization accuracy and efficiency by referencing the colorization model with a simplified color palette.

—A customized self-guided automatic colorization network IsPlanktonCLR is designed and its colorization performance on plankton images has been verified superior to SOTA methods. To the best of our knowledge, this is the first endeavor to study scientific colorization of *in situ* marine plankton imagery.

—A new metric CDSIM is proposed for evaluation of color similarity between input and output images of a colorization model, which has been verified suitable for restoration-oriented image colorization problems.

## 2 Related Work

### 2.1 Underwater plankton cameras for *in situ* plankton color imaging

All the underwater cameras that are good at capturing *in situ* colorful images of marine plankton have adopted dark-field imaging principle. The early Video Plankton Recorder (VPR) can only capture grayscale images [9], but it is reported to support color imaging latterly after device upgrade [33]. The Continuous Particle Imaging and Classification System (CPICS) has been deployed on the sea floor [14]. The Scripps Plankton Camera (SPC) has been deployed underwater at the shore [31], with profilers [6], and under a floating station in the Lake Greifensee [28]. The Imaging Plankton Probe (IPP) has been deployed under a moored buoy for monitoring coastal waters [21].

There is no significant difference in the imaging light path of these dark-field cameras. While they all use flashed white-light as sources, their lighting path design is different. VPR uses lateral side lighting [9], CPICS uses annular oblique lighting [13], SPC uses standard hollow-cone illumination as adopted in traditional dark-field microscopes [31], and IPP uses annular orthogonal compressed lighting [21]. VPR and SPC are not reported to be color-calibrated, while CPICS and IPP both perform white balance calibration using external reference targets [13,21]. It is worth noting that IPP has not only achieved high-quality *in situ* true color imaging of marine plankton through spatially compressed and condensed laminar white-light illumination, but also greatly inhibited the leakage of white light to the adjacent underwater environment, thus greatly reducing the phototropic aggregation of zooplankton [21]. However, in principle, this lighting design still cannot completely eliminate the white-light leakage scattered by the seawater within the imaging area. It is still suspicious whether the influence by phototaxis of zooplankton can be completely avoided.

### 2.2 Deep learning-based image colorization

The colorization of grayscale image has long been a very challenging problem. With the development of deep learning technology and the emergence of large-scale image datasets such as ImageNet [10] and MSCoco [24], various deep networks have been applied to the field of natural scene image colorization and

achieved good results [1,45,44,20,41,40]. Based on the difference of their objectives, these deep image colorization algorithms can be roughly classified into two groups for image enhancement and restoration. The algorithms for image enhancement [8,17,41,40,34,46,27,3,42] aim at converting grayscale images into color images with visual comfortableness and fit of human commonsense, but pay little attention to whether the generated colors are the same as ground truth.

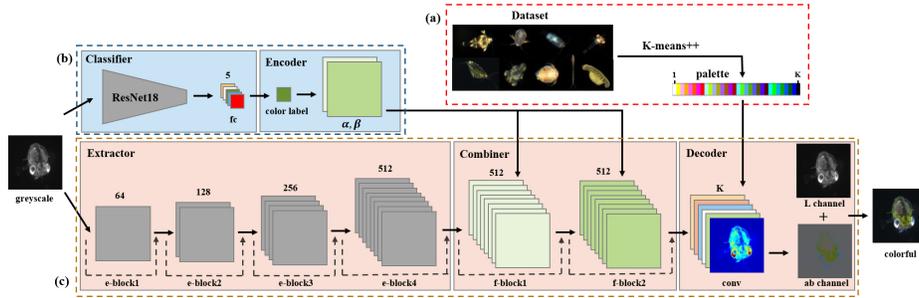
The deep colorization algorithms for image restoration aim to recover the original true color of scenes or objects in the grayscale images. CIC [45] transforms the colorization problem into a classification problem of 313 colors. Pixelated Semantic Colorization uses pixel-level semantic features to guide a network for colorization [47]. LetColor achieves end-to-end colorization by learning the global prior and local features of the image [19]. MemoColor [44] remembers color information of rare instances through an external storage network [20], achieving colorization with limited data. We notice some networks use paired image data for colorization learning. For example, Colorization in the dual-lens system [11] and Low-light Color Imaging [15] both use dual-camera systems to obtain grayscale-color image pairs at the same time. Then the grayscale images are colorized with the guidance from their color counterparts, so that the color information is transferred from the color images with low resolution and poor brightness to the grayscale images with higher quality. This is similar to our idea, except that their goal is to improve the resultant quality of natural scene color images, while ours is to automatically and accurately restore the true color of marine plankton in grayscale images to be taken under red-NIR lighting in seawater.

### 2.3 Metrics for colorization evaluation

Existing metrics for image colorization are mainly used to evaluate whether the color of network output images conforms to human commonsense and cognition. For example, the Colorfulness Score [16] evaluates the quality and diversity of image colorization, but it is difficult to compare the color authenticity of the output image with its ground truth. UCIQE [43] evaluates the color quality of underwater images through a linear combination of chroma, saturation and contrast, but it does not consider the spatial distribution of color. Although PSNR and SSIM are also frequently used for image restoration evaluation, they are proved insensitive to color difference between images [17,40,47,7,29,37]. FID [18] calculates the Fréchet distance between feature vectors extracted by an Inception V3 network to evaluate the overall similarity between two comparative images, but the deep features are not interpretable enough to clarify the relationship between both color images. In a word, it is difficult for previous metrics to simulate the perception of human vision, and to objectively and quantitatively evaluate the color similarity between images.

### 3 Methodology

As illustrated in Figure 2, the IsPlanktonCLR network mainly consists of two parts: a customized color palette, which simplifies the color space of the plankton image data; and a deep colorization network, which is used to achieve self-guided and automated colorization of the input grayscale image.



**Fig. 2.** Overview of IsPlanktonCLR: (a) palette customization; (b) reference module; (c) colorization module.

#### 3.1 Palette customization

We use the color ROI images in the DYB-PlanktonNet [22] dataset to customize the reference palette. For convenience, we convert all the RGB ROIs into *Lab* color space with  $(a_i, b_i)$  denotes the  $i^{th}$  color. Since the background of dark-field image is nearly zero, we only extract color information from the foreground pixels and integrate similar colors with a clustering algorithm based on *K*-means++ [2] to reduce the number of colors. To achieve this, we firstly select manually 2 ~ 3 ROIs from each plankton class of the dataset, which can represent as more colors as possible in this class. Then the foreground pixels of these ROIs are extracted by thresholding, and their  $(a, b)$  values are used to train the *K*-means++ clustering algorithm. After 10 iterations, the model with the smallest inertia (sum of squared distances of samples to their closest cluster center) is selected and denoted as  $Model_{km}$ . The palette is then customized to contain colors represented by the clustering centers of  $Model_{km}$ :

$$palette = \{(a_i, b_i), i \in K\}, \quad (1)$$

where  $K$  is the number of clustering centers. We select a reasonable value range of  $K$  by observing the inertia variation curve with  $K$ , and finally determine an optimized value of  $K$  by comparing the image colorization results obtained with different  $K$ .

### 3.2 Colorization Network

As shown in Fig. 2 (b) and (c), the IsPlanktonCLR network consists of two parallel modules. We classify the plankton images in the DYB-PlanktonNet into five color families, namely white, red, yellow, green, and blue. The reference module firstly determines the color family labels of an input grayscale image, and then the colorization module completes the colorization under the guidance of these labels.

The reference module consists of a classifier and an encoder. The classifier is built on a ResNet18 network and is responsible for giving a color family label  $L$  to the input grayscale image  $G$ .  $L$  is an integer ranging from 1 to 5, corresponding to 5 color family labels, respectively. The encoder is used to encode the color family label  $L$  and generate the reference information  $\alpha$  and  $\beta$  for the colorization module. We test two encoding methods. The first one is discrete encoding, which directly encodes  $L$  with embedding to obtain the reference information  $R$ ; the other is continuous encoding, whose formula is

$$R = L - 1 + P, \quad (2)$$

where  $P$  is the probability that  $G$  belongs to  $L$ . The  $R$  values obtained by the two encoding methods are fed into an  $1 \times 1$  convolution layer, and finally split into  $\alpha$  and  $\beta$ .

The colorization module is mainly composed of an extractor, a combiner and a decoder. The extractor is mainly responsible for extracting the feature map  $fm_{ex}$ , which consists of four e-blocks. Each e-block contains three convolutional layers with batch-normalization and ReLU. The combiner is responsible for adding  $\alpha$  and  $\beta$  to  $fm_{ex}$ , which consists of two f-blocks. Each f-block is composed of convolutional and FILM layers [27]. FILM layer implements the function as expressed in Equation (3), where  $fm_{ref}$  is a feature map with reference information.

$$fm_{ref} = \alpha fm_{ex} + \beta. \quad (3)$$

The decoder achieves color classification of each pixel by the convolutional and softmax layers. We encode the softmax output as a one-hot vector and then decode it by multiplying it with the palette vector to obtain the  $ab$  channel of the image. Finally, the  $ab$  channel is overlapped with the input  $L$  channel to form the colorized image.

### 3.3 Loss Function

We treat pixels containing common colors as easy examples, and pixels containing rare colors as hard examples. We use a loss function incorporating OHEM [35] and Focal-Loss [23] to solve the color imbalance problem by training the model with more weights on rare colors to avoid dominant color issue in the results. The loss function is formulated as follows:

$$loss_{pixle}(x, p_t) = -\omega_t (1 - p_t)^\gamma \log(p_t), \quad (4)$$

$$loss = \frac{1}{N} \sum_{i \in S} loss_{pixle}(i, p_t). \quad (5)$$

In Formula (4),  $loss_{pixle}(x)$  denotes the loss of classifying pixel  $x$ , where  $p_t$  denotes the probability that  $x$  is classified correctly, and  $\omega_t$  denotes the weight of the correctly classified category in calculating the loss.  $\omega$  is determined by  $Model_{km}$  in Section 3.1. We use  $Model_{km}$  to quantify the image color in the entire training set to obtain information about the number of colors in each class. The higher the number of colors, the smaller the weight is assigned to pixels belong to that color class. Therefore,  $\omega$  is mainly used to balance the differences between categories.  $\gamma$  is a modulating factor to make the weights of hard examples larger than those of easy examples for balance.

In Formula (5),  $S$  denotes the set of hard examples, and  $N$  is the number of hard examples. We first sort the color classification loss of all pixels in a batch, and then take  $N$  pixels with the largest loss to form  $S$ . This allows the model to focus only on the pixels with larger loss and ignore the remaining pixels. In addition, as the training progresses, the number of hard examples gradually decreases, so the size of  $N$  also decreases.

The reference module and the colorization module produces a loss as expressed by Formula (5), respectively. We add the two in proportion as the final training loss as follows:

$$loss_{final} = \mu loss_{ref} + \vartheta loss_{color}, \quad (6)$$

where  $\mu$  and  $\vartheta$  are the scaling factors for the reference module loss  $loss_{ref}$  and colorization module loss  $loss_{color}$ , respectively.

### 3.4 Evaluation Metric

We propose CDSIM to evaluate the difference in color quantity and spatial distribution between the colorization results and their ground truth. The calculation of this metric is dependent on the extraction of color features from the images. The color features we use include color histogram, color coherence vector, color correlogram, and color gradients. Details of their definition and feature reduction can be referred to the Supplementary Materials.

The combination and reduction of these features can produce an 1260-dimensional vector. Then the Euclidean distance between the color feature vectors of two images is defined as CDSIM, which is expressed in Equation (7).

$$CDSIM(X, Y) = \sqrt{\sum_{i=1}^l (x_i - y_i)^2}, \quad (7)$$

where  $X$  and  $Y$  are the color feature vectors of two comparative images, and  $l$  is the vector length. The smaller the CDSIM is, the better the colorization is.

## 4 Experiments

### 4.1 Dataset

We construct a dataset consisting of 2907 *in situ* marine plankton ROI images for training and testing the IsPlanktonCLR network, whose composition is shown in Table 1. There are two main sources for this dataset. The first is the DYB-PlanktonNet dataset [22], which contains ROI images of 92 classes of marine plankton and suspended particles recorded *in situ* by IPP [21]. The second is from a dataset we exclusively obtained for this study by imaging natural seawater sample with the customized dual-channel dark-field imaging apparatus as used in [26]. For acquiring the IsPlanktonCLR dataset, this apparatus is firstly modified with installment of two lenses with the same magnification, replacement of one color camera with a grayscale camera with the same pixel size and number, and addition of an 850nm NIR light source to the white light source to illuminate the plankton in the seawater sample. Then the grayscale and color cameras are synchronized to capture image pairs of the same plankton sample in real seawater, which eventually constitute the IsPlanktonCLR dataset after image registration similar to that used in [26]. The IsPlanktonCLR dataset is available at [https://drive.google.com/drive/folders/1GspuXRqd\\_GbB2k12Uwic1N3MPFoclxYn?usp=sharing](https://drive.google.com/drive/folders/1GspuXRqd_GbB2k12Uwic1N3MPFoclxYn?usp=sharing).

**Table 1.** Composition of the *in situ* plankton image dataset for experiments.

	Training Set	Testing Set 1	Testing Set 2	Total
DYB-PlanktonNet	2117	356	0	2473
IsPlanktonCLR	344	30	60×2	494
Total	2461	386	120	2967

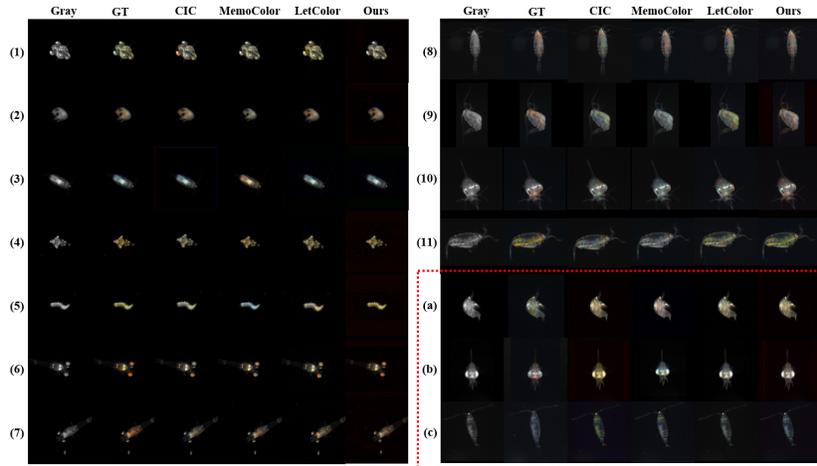
### 4.2 Comparisons with Previous Works

Figure 3 compares the colorization results of multiple marine plankton grayscale images produced by IsPlanktonCLR and several SOTA approaches. Among them, images in row (1) ~ (11) are results from Testing Set 1, and images in row (a) ~ (c) are results from Testing Set 2.

Judging from human visual perception, the colorization performance of IsPlanktonCLR is obviously better than other models. The plankton shown in row (1) ~ (4) are a Megalopa larvae, an amphipod, and two copepods, which are numerous and common in the dataset. The rest of the tested images contain rare colors, on which the results of IsPlanktonCLR remained excellent, but the results of other models on them are significantly degraded. Specifically, other models show varying degrees of dominant color effect in row (5) ~ (7), color averaging

effect in row (8), wrong colorization in row (9) ~ (10), and poor colorization of details in row (11).

IsPlanktonCLR also achieves good colorization effect on the images from Testing Set 2, *i.e.*, row (a) ~ (c), while the colorization results of other models are far from the ground truth. This demonstrates very good device- and content-generalization potential of the IsPlanktonCLR network.



**Fig. 3.** Visual perception comparison of colorization performance by IsPlanktonCLR and other SOTA methods.

Table 2 compares the evaluation results between IsPlanktonCLR and other models on various numerical metrics. IsPlanktonCLR obtains the highest scores on both CDSIM and FID, which proves that the images generated by this algorithm have the highest similarity with ground truth; it performs slightly worse than other models on Colorfulness Score, because this metric mainly measures the color richness of image and is irrelevant to the colorization accuracy. Although the results of MemoColor get the best evaluation on Colorfulness Score, they contain many unreal colors. We also provide the evaluation results of PSNR and SSIM for reference only.

In order to make a fair comparison between IsPlanktonCLR and SOTA approaches, we further conduct an online survey to collect human visual evaluation from 115 volunteers (mainly composed of PhD students, marine biologists and several marine plankton experts) on the color similarity between the colorization results of four models and the ground truth. The survey questionnaire is designed to include 14 groups of marine plankton images, which can be referred to in the Supplemental Materials. The volunteers are asked to score the color similarity between each colorized image generated by one of four colorization models with its ground truth. The score is based on an 1-5 points scale with

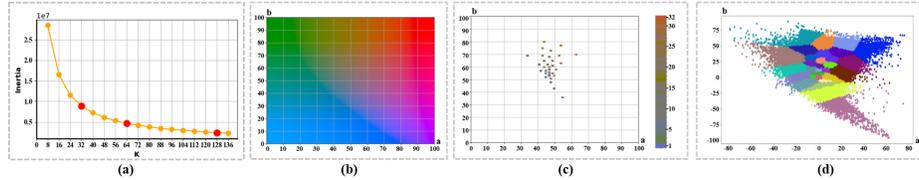
5 for the most similar and 1 for the least similar. The average score for each model is finally tabulated in the rightmost column in Table 2, which indicates that IsPlanktonCLR still performs the best.

**Table 2.** Numerical comparison of colorization performance by IsPlanktonCLR and other SOTA methods under various evaluation metrics.

	CDSIM ↓	FID ↓	Colorfulness ↑	UCIQE ↑	PSNR ↑	SSIM ↑	Human ↑
CIC	466.001	40.904	5.901	0.443	42.903	<b>0.997</b>	2.716
MemoColor	734.186	29.348	<b>6.789</b>	0.429	43.307	0.995	2.668
LetColor	384.637	29.063	6.009	<b>0.447</b>	43.905	0.983	2.920
IsPlanktonCLR	<b>346.434</b>	<b>24.578</b>	5.921	0.425	<b>44.269</b>	0.996	<b>3.785</b>

### 4.3 Ablation Experiments

**Color Number.** In order to select an appropriate number of color clusters  $K$  during palette customization, we calculate the variation of inertia with the number of clusters  $K$  as shown in Figure 4 (a). In general, the smaller the inertia, the better the color clustering, and the larger the corresponding  $K$ . We compare the colorization results when  $K$  is taken as 32, 64, and 128, respectively, and find that there is little difference among them. Therefore, we choose  $K = 32$  as the final number of clusters.



**Fig. 4.** (a) Inertia variation with the color cluster numbers. (b) The palette of the natural scene, (fixing  $L = 50$  and normalizing  $ab$  channel values to  $[0,100]$ ). (c) The customized palette of DYB-PlanktonNet dataset at  $K = 32$ . (d) Color clustering result of the DYB-PlanktonNet dataset with the customized palette (colors are not real but only for visualization).

As can be compared in Figure 4 (b) and (c), the customized palette of the *in situ* plankton images has significantly reduced colors over the palette of natural scenes after color clustering. This not only greatly simplifies the search space of

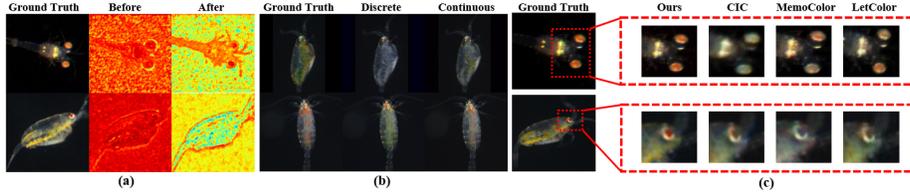
the colorization algorithm, but also limits the color abuse by the model. Figure 4 (d) shows the 32 clustered colors of the DYB-PlanktonNet dataset are well separated.

**Color label.** To validate the effectiveness of reference information provided by the reference module, the activation maps before and after the combiner in the colorization module are visualized in Figure 5 (a). It can be seen that before the reference information is added, the network can only distinguish foreground plankton from the background, and the mean activation within the plankton is relatively homogeneous. After the addition of the reference information, the mean activation at different colors within the plankton appears significantly different. We can see the values of reddish colors in the ground truths are higher in the activation map, while the those of yellowish or greenish colors are lower. These results prove that the reference information can really help the network to distinguish different colors and provide effective guidance for colorization

We also compare the colorization effect between discrete encoding and continuous encoding. For the majority of images, there is little difference between the two encoding methods. However, for some plankton with similar morphology and are difficult to determine color family labels, the continuous encoding has achieved slightly better colorization effects than those obtained by discrete encoding. Figure 5(b) shows the difference of colorization effect between the two encoding methods on two confusable examples. It can be seen that the colorization module will generate wrong colors with the guidance of discrete encoding when the reference module gives a wrong color family label, while the continuous coding can enable the model to recover correct colors in many positions. Taking the bottom row of Figure 5(b) as an example, the network wrongly generates the greenish colors with discrete encoding, while it can recover correctly reddish colors with continuous encoding.

**Loss Function.** During training, we set the initial number of  $N$ , *i.e.*, the difficult examples, to be 0.05 of the total number of pixels, and it decays every 200 epochs with an attenuation rate of 0.5. Both the reference module and the colorization module use the same loss function as expressed in Formula (6) with coefficients  $\mu$  and  $\vartheta$  equal to 0.1 and 0.9, respectively.

Figure 5(c) compares the effects of IsPlanktonCLR and other models on colorization of rare colors. As the bodies of most plankton are semi-transparent or transparent, there are many white examples in our dataset while the reddish examples are rare. Except for a few species, reddish colors mainly appear in the positions of plankton eyes. As can be seen, our model overcomes the dominant color effect well and achieves accurate colorization of the decapod’s eyes and the copepod’s eye-spot, while other models perform much poorer. In addition, we conduct further experiment to compare the performance of our loss with the cross-entropy loss as baseline. The results are detailed in the Supplementary Materials, which show that our loss can make the model converge faster than the baseline loss does under the same conditions.



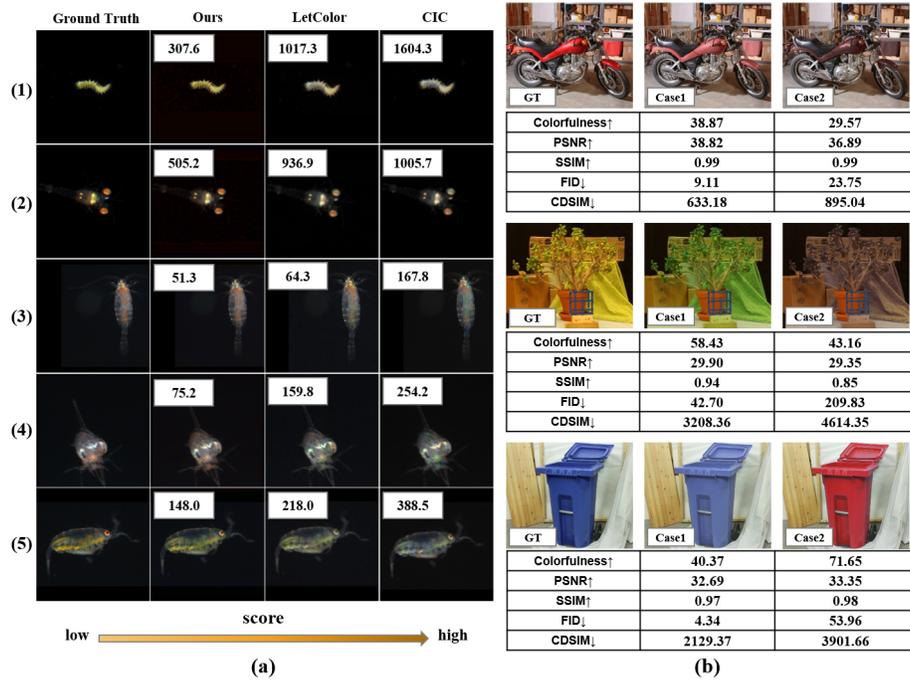
**Fig. 5.** (a) Examples indicating the change in mean activation before and after the addition of reference information. The colors in the activation maps are related to the mean value of activation with low values indicating cooler colors and high values indicating warmer colors. (b) Comparison of colorization effects between discrete encoding and continuous encoding. (c) Colorization effect comparison of each model on rare colors.

#### 4.4 CDSIM Metric

We firstly use a dataset of *in situ* plankton images to validate the proposed CDSIM. Figure 6 (a) shows the evaluation results of color dissimilarity obtained by different colorization models, where only the foreground pixels of plankton are used for CDSIM calculation. Intuitively, the color differences of images in each row from the ground truth become gradually more obvious from left to right, and their CDSIM scores also gradually increase, indicating that the colorization quality is getting worse. This result verifies that CDSIM can not only quantify color dissimilarity between plankton images, but also its evaluation results are consistent with the perception of human eyes.

In addition, we select some images from ImageNet [10] to further assess CDSIM on natural scene image colorization, and the results are compared with those obtained by other common metrics, as shown in Figure 6(b). Before this test, we replace some colors in the original pictures with other colors that look still reasonable to human eyes. In Case1, some colors of the scenes are replaced with visually similar colors; while in Case2, some colors in the pictures are replaced with obviously different colors. In this test, all the pixels of a whole image are evaluated. The results show that when using CDSIM and FID, the scores of Case1 images are significantly lower than those of Case2, which is consistent with human visual perception. But the results of PSNR and SSIM are very similar in the two cases, indicating that they cannot distinguish the color difference. The evaluation results of Colorfulness are not consistent at all. This result proves that CDSIM is also suitable for objective and quantitative evaluation of natural scene image colorization.

Compared with PSNR, SSIM, FID and other metrics, CDSIM has higher computational complexity. Especially for high resolution natural scene colorful images, its computation can be intensive. However, for *in situ* marine plankton dark-field ROIs, the CDSIM computation cost is significantly reduced due to limited image resolution and even lower proportion of foreground pixels, which is completely acceptable in practice.



**Fig. 6.** (a) Comparison of visual perception and CDSIM evaluations on marine plankton images produced by various colorization models. (b) Comparison of visual perception and numerical metrics-based evaluations on artificially colored natural scene images.

## 5 Conclusion

We present a deep colorization model for automatic color restoration of *in situ* marine plankton grayscale images. The model achieves the state-of-the-art performances on color restoration of *in situ* marine plankton image data. This is the first endeavor, to the best of our knowledge, to apply deep colorization for marine plankton scientific imagery. We also propose a metric for comparing color dissimilarity between images, which provides a new and objective evaluation for restoration-oriented image colorization algorithms. This method is expected to inspire new design of next generation instruments or systems for achieving long-term, continuous, high-frequency, and *in situ* ocean observation.

**Acknowledgements.** This work was supported by International Partnership Program of Chinese Academy of Sciences No.172644kysb20210022, Scientific Instrument Development Project of Chinese Academy of Sciences No.YJKYYQ20190028, and Shenzhen Science and Technology Innovation Program No.JCYJ20200109105823170. We thank the participants for replying our online survey.

## References

1. Anwar, S., Tahir, M., Li, C., Mian, A., Khan, F.S., Muzaffar, A.W.: Image colorization: A survey and dataset. arXiv preprint arXiv:2008.10774 (2020)
2. Arthur, D., Vassilvitskii, S.: k-means++: The advantages of careful seeding. Tech. rep., Stanford (2006)
3. Bahng, H., Yoo, S., Cho, W., Park, D.K., Wu, Z., Ma, X., Choo, J.: Coloring with words: Guiding image colorization through text-based palette generation. In: Proceedings of the european conference on computer vision (eccv). pp. 431–447 (2018)
4. Baig, M.H., Torresani, L.: Multiple hypothesis colorization and its application to image compression. *Computer Vision and Image Understanding* **164**, 111–123 (2017)
5. Benfield, M.C., Grosjean, P., Culverhouse, P.F., Irigoien, X., Sieracki, M.E., Lopez-Urrutia, A., Dam, H.G., Hu, Q., Davis, C.S., Hansen, A., et al.: Rapid: research on automated plankton identification. *Oceanography* **20**(2), 172–187 (2007)
6. Campbell, R., Roberts, P., Jaffe, J.: The prince william sound plankton camera: a profiling in situ observatory of plankton and particulates. *ICES Journal of Marine Science* **77**(4), 1440–1455 (2020)
7. Cao, Y., Zhou, Z., Zhang, W., Yu, Y.: Unsupervised diverse colorization via generative adversarial networks. In: Joint European conference on machine learning and knowledge discovery in databases. pp. 151–166. Springer (2017)
8. Ci, Y., Ma, X., Wang, Z., Li, H., Luo, Z.: User-guided deep anime line art colorization with conditional adversarial networks. In: Proceedings of the 26th ACM international conference on Multimedia. pp. 1536–1544 (2018)
9. Davis, C., Gallager, S., Berman, M., Haury, L., Strickler, J.: The video plankton recorder (vpr): design and initial results. *Arch. Hydrobiol. Beih* **36**, 67–81 (1992)
10. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
11. Dong, X., Li, W.: Shoot high-quality color images using dual-lens system with monochrome and color cameras. *Neurocomputing* **352**, 22–32 (2019)
12. Forward, R.B.: Light and diurnal vertical migration: photobehavior and photophysiology of plankton. In: Photochemical and photobiological reviews, pp. 157–209. Springer (1976)
13. Gallager, S.M.: Continuous particle imaging and classification system (Mar 5 2019), uS Patent 10,222,688
14. Grossmann, M.M., Gallager, S.M., Mitarai, S.: Continuous monitoring of near-bottom mesoplankton communities in the east china sea during a series of typhoons. *Journal of oceanography* **71**(1), 115–124 (2015)
15. Guo, P., Ma, Z.: Low-light color imaging via dual camera acquisition. In: Proceedings of the Asian Conference on Computer Vision (2020)
16. Hasler, D., Suesstrunk, S.E.: Measuring colorfulness in natural images. In: Human vision and electronic imaging VIII. vol. 5007, pp. 87–95. International Society for Optics and Photonics (2003)
17. He, M., Chen, D., Liao, J., Sander, P.V., Yuan, L.: Deep exemplar-based colorization. *ACM Transactions on Graphics (TOG)* **37**(4), 1–16 (2018)
18. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. *Advances in neural information processing systems* **30** (2017)

19. 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)
20. Kaiser, L., Nachum, O., Roy, A., Bengio, S.: Learning to remember rare events. arXiv preprint arXiv:1703.03129 (2017)
21. 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)
22. Li, J., Yang, Z., Chen, T.: Dyb-planktonnet. *IEEE Dataport* (2021)
23. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollár, P.: Focal loss for dense object detection. In: *Proceedings of the IEEE international conference on computer vision*. pp. 2980–2988 (2017)
24. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context. In: *European conference on computer vision*. pp. 740–755. Springer (2014)
25. Lombard, F., Boss, E., Waite, A.M., Vogt, M., Uitz, J., Stemann, L., Sosik, H.M., Schulz, J., Romagnan, J.B., Picheral, M., et al.: Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science* p. 196 (2019)
26. Ma, W., Chen, T., Zhang, Z., Yang, Z., Dong, C., Qiao, J., Li, J.: Super-resolution for in situ plankton images. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 3683–3692 (2021)
27. Manjunatha, V., Iyler, M., Boyd-Graber, J., Davis, L.: Learning to color from language. arXiv preprint arXiv:1804.06026 (2018)
28. Merz, E., Kozakiewicz, T., Reyes, M., Ebi, C., Isles, P., Baity-Jesi, M., Roberts, P., Jaffe, J.S., Dennis, S.R., Hardeman, T., et al.: Underwater dual-magnification imaging for automated lake plankton monitoring. *Water Research* **203**, 117524 (2021)
29. Messaoud, S., Forsyth, D., Schwing, A.G.: Structural consistency and controllability for diverse colorization. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. pp. 596–612 (2018)
30. Nazeri, K., Ng, E., Ebrahimi, M.: Image colorization using generative adversarial networks. In: *International conference on articulated motion and deformable objects*. pp. 85–94. Springer (2018)
31. Orenstein, E.C., Ratelle, D., Briseño-Avena, C., Carter, M.L., Franks, P.J., Jaffe, J.S., Roberts, P.L.: The scripps plankton camera system: A framework and platform for in situ microscopy. *Limnology and Oceanography: Methods* **18**(11), 681–695 (2020)
32. Picheral, M., Grisoni, J.M., Stemann, L., Gorsky, G.: Underwater video profiler for the” in situ” study of suspended particulate matter. In: *IEEE Oceanic Engineering Society. OCEANS’98. Conference Proceedings (Cat. No. 98CH36259)*. vol. 1, pp. 171–173. IEEE (1998)
33. Plonus, R.M., Conradt, J., Harmer, A., Janßen, S., Floeter, J.: Automatic plankton image classification—can capsules and filters help cope with data set shift? *Limnology and Oceanography: Methods* **19**(3), 176–195 (2021)
34. Sangkloy, P., Lu, J., Fang, C., Yu, F., Hays, J.: Scribbler: Controlling deep image synthesis with sketch and color. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 5400–5409 (2017)
35. Shrivastava, A., Gupta, A., Girshick, R.: Training region-based object detectors with online hard example mining. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 761–769 (2016)

36. Steinberg, D.K., Landry, M.R.: Zooplankton and the ocean carbon cycle. *Annual review of marine science* **9**, 413–444 (2017)
37. 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)
38. Tanaka, M., Genin, A., Endo, Y., Ivey, G.N., Yamazaki, H.: The potential role of turbulence in modulating the migration of demersal zooplankton. *Limnology and Oceanography* **66**(3), 855–864 (2021)
39. Tanaka, M., Genin, A., Lopes, R.M., Strickler, J.R., Yamazaki, H.: Biased measurements by stationary turbidity-fluorescence instruments due to phototactic zooplankton behavior. *Limnology and Oceanography: Methods* **17**(9), 505–513 (2019)
40. 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)
41. Wu, Y., Wang, X., Li, Y., Zhang, H., Zhao, X., Shan, Y.: Towards vivid and diverse image colorization with generative color prior. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 14377–14386 (2021)
42. Xu, Z., Wang, T., Fang, F., Sheng, Y., Zhang, G.: Stylization-based architecture for fast deep exemplar colorization. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. pp. 9363–9372 (2020)
43. Yang, M., Sowmya, A.: An underwater color image quality evaluation metric. *IEEE Transactions on Image Processing* **24**(12), 6062–6071 (2015)
44. 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)
45. Zhang, R., Isola, P., Efros, A.A.: Colorful image colorization. In: *European conference on computer vision*. pp. 649–666. Springer (2016)
46. Zhang, R., Zhu, J.Y., Isola, P., Geng, X., Lin, A.S., Yu, T., Efros, A.A.: Real-time user-guided image colorization with learned deep priors. *arXiv preprint arXiv:1705.02999* (2017)
47. Zhao, J., Han, J., Shao, L., Snoek, C.G.: Pixelated semantic colorization. *International Journal of Computer Vision* **128**(4), 818–834 (2020)