

Super-resolution for *in situ* Plankton Images

Wenqi Ma¹ Tao Chen^{2,3} Zhengwen Zhang² Zhenyu Yang^{2,3} Chao Dong²
 Jianping Qiao¹ Jianping Li^{2,3,†}

¹Shandong Normal University, China

²Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, China

³University of Chinese Academy of Sciences, China

†jp.li@siat.ac.cn

Abstract

Being inherently limited by the wave properties of light, underwater plankton cameras compromise between their imaging resolution and field of view (FOV) for *in situ* observations. In order to enlarge the sampling volume in single frame acquisition, lower magnifications are usually adopted to enable larger FOV but sacrifice the resolution. In this paper, we build a real-underwater image dataset called IsPlanktonSR for *in situ* plankton image super-resolution (SR), in which paired low resolution (LR) and high resolution (HR) images of the same individual live planktonic organisms are captured by a customized dual-channel darkfield imaging system. An image registration algorithmic pipeline is also proposed to preprocess and align the image pairs at different scaling factors of $2\times$ and $4\times$. The IsPlanktonSR dataset is used to train an enhanced deep residual network for SR through the L2, the perceptual and the contextual losses, respectively. Our extensive experimental results demonstrate that the deep learning model trained on real data through the contextual loss has delivered better visual and quantitative SR performance than those trained on simulated data or through other loss functions. The trained SR model is also proved to generalize well to images of various plankton species or captured by different instruments. The proposed SR technology is anticipated to enhance the existing darkfield plankton imageries and enable the future *in situ* plankton imaging instruments for better observation capability and hence deepen understanding of the plankton ecology.

1. Introduction

In situ optical imaging technology can capture images of the marine plankton in their natural state in seawater, and has become a new powerful means to study the marine plan-

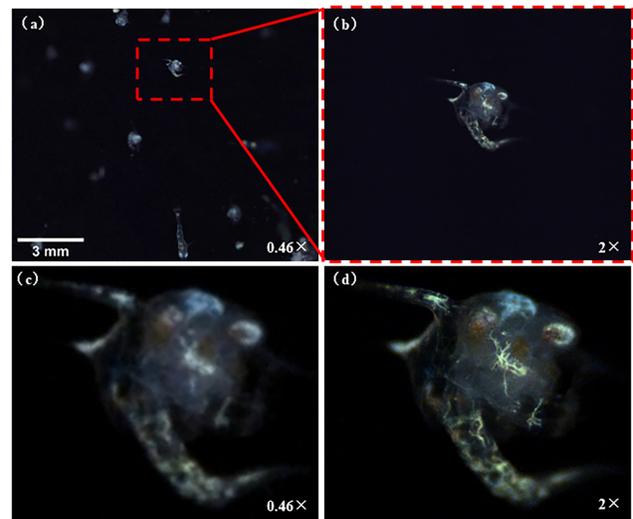


Figure 1. Compromise between magnification (FOV) and resolution in underwater plankton imaging. (a) and (b) are darkfield plankton images at magnifications of $0.46\times$ and $2\times$; (c) and (d) are aligned real LR-HR underwater plankton image pair after image registration. (c) is displayed after bicubic interpolation for a side-by-side comparison with (d).

tonic ecosystem [22]. Since 1990s, several categories of *in situ* optical imaging systems have been developed for plankton observation. Generally, they can be classified into the brightfield and the darkfield imagers. Brightfield imagers are often deployed in mobile platforms such as research vessel [1, 7, 8, 23, 33] and even autonomous glider [28]. These imagers are usually poorer in imaging resolution and contrast [1, 7], and their images are susceptible to interference from the excessive debris in coastal waters [6, 10]. Darkfield cameras generally have higher resolution and contrast, and are more suitable for long-term high-frequency continuous observations at fixed spots, such as the shore [29], the sea floor [11], and even under a surface buoy [18].

There is a serious issue faced with existing imaging systems. Specifically, in real scenarios, the density of plankton is often sparse in seawater. In order to observe more plankton individuals, existing *in situ* optical imaging systems have to sample more water volume per unit time, resulting in a sacrifice of imaging magnification. Note that the sampling volume per image is usually estimated by the product of the field of view (FOV) and the depth of field (DOF). However, this will inevitably lead to a decreased imaging resolution, which is insufficient to obtain enough image details for the relatively small plankton targets, as shown in Figure 1. Besides, this phenomenon will also seriously affect the accuracy of subsequent plankton taxonomic recognition and quantification.

Image super-resolution (SR) refers to the image restoration technology that recovers a high-resolution (HR) image from its low-resolution (LR) counterpart [38]. In recent years, deep learning based SR has made substantial progress in various imaging modalities, such as the natural scene images [15, 17, 20], medical images [12, 13], light microscopy images [32, 39] and so on. If the SR technology could be used to improve the resolution of the *in situ* plankton images on the premise of reserving their original FOV and DOF, it would be helpful to obtain indistinguishable details of the small organisms, and possibly improve the accuracy of subsequent species identification [3].

Most of the existing training datasets for perceptual SR are constructed by subsampling HR images. However, these methods fail to super-resolve real-world LR images, because the degradation process is much more complicated and unknown [24]. In recent years, a few real-world SR datasets have been proposed, e.g., the RealSR [4] and the City100 [5], which have improved the SR performance for the DSLR and phone cameras. Since the camera systems for *in situ* plankton imaging have to work in more harsh and complex underwater environments than terrestrial settings, the degradation model between their HR and the LR images is even more complex and unpredictable. *In situ* plankton imaging belongs to macro/microscopic photography, and the speed of plankton swimming (or with the water flow) is very fast. These facts will make pixel-level image registration very difficult when constructing the real-underwater image pairs, and the lack of registration accuracy will seriously affect the performance of the trained SR models [39]. Therefore, direct training SR networks by the simulated data or by existing real SR datasets is unlikely to produce satisfactory results for underwater plankton image SR. It is necessary to construct a real-underwater plankton dataset. However, it is by no means a trivial task.

In order to address the above problem, we design a dual-channel darkfield imaging system that is capable of simultaneously capturing HR-LR image pairs of the same individual live plankton organisms in a real underwater environ-

ment. We capture a large number of raw HR-LR image pairs of living plankton in the first place by using this system. Then we apply a series of image processing techniques, including image correction, target detection, focus evaluation, and image registration, to generate a real-underwater plankton image dataset, called IsPlanktonSR (Is stands for *in situ*). Further, we train an Enhanced Deep Residual Network (EDSR) [20] with the IsPlanktonSR dataset, and compare its SR performance on different training datasets under the L2 loss, the perceptual loss [14], and the contextual loss [25], respectively. After we have confirmed that the combination of using real datasets and contextual loss performs the best, we train a $2\times$ and a $4\times$ SR model with the IsPlanktonSR dataset. Finally, we apply the trained $4\times$ SR model on images captured by various *in situ* imaging systems and a laboratory stereoscopic microscope to test the generalizability of the method. Extensive experiments demonstrate that all these tests have delivered good results.

In summary, the major contributions of this work are:

- We build a real-underwater IsPlanktonSR dataset consisting of registered HR-LR plankton image pairs with $2\times$ and $4\times$ scaling factors, providing an *in situ* marine plankton image benchmark for real-underwater SR model training and evaluation.
- We experimentally verify that training the EDSR model with contextual loss on the IsPlanktonSR dataset can achieve satisfying SR effect for *in situ* darkfield plankton images.
- The developed SR algorithm can improve the image resolution of existing darkfield plankton image data, and provided new design possibilities for developing future multi-resolution underwater plankton imagers.

2. Related Work

***In situ* Plankton Darkfield Imaging Systems.** Among the underwater darkfield cameras developed for *in situ* plankton observations, the Scripps Plankton Cameras (SPC) [29], continuous plankton imaging and classification system (CPICS) [11], and the underwater darkfield plankton imager developed by Li *et al.* [18] all used a strategy of supporting installation and replacement of one telecentric lenses with different magnifications. Thus, each magnification can support the observation of the plankton groups within a certain size range. Obviously, this strategy cannot enable simultaneous observation of plankton at different resolutions and the replacement of different magnification lenses is troublesome and costly. Recently, the SPC have evolved a new version to support two magnifications of $5\times$ and $0.5\times$ lenses in the same housing for simultaneous observation of both zooplankton and phytoplankton [26]. This strategy is equivalent to installing two cameras in one housing and hence raises the cost and system complexity.

Wang *et al.* devised an underwater darkfield camera using a motorized lens nosepiece supporting the rotary switching among three lenses with different magnifications [36]. Such a scheme is not only expensive and bulky, but also increases the potential risk of system failure when applied in the fields for long-term deployment. From the hardware perspective, there is currently a lack of underwater imaging systems that can support different magnification and resolution for *in situ* plankton observations.

Deep Learning-based SR. Since Dong *et al.* proposed the SRCNN [9], many learning-based SR models using HR-downsampled image pair datasets for training and testing have emerged, and their SR performance have also been continuously improved [40, 21, 37]. However, the mapping between the real HR-LR image pair is often more complex and unpredictable than that can be extracted from simulated data training. As a result, the SR performance of these efforts are often unsatisfactory for real-world LR images.

To further improve the SR performance, people have tried to build real-world datasets by developing customized imaging system and image registration algorithms to produce spatially aligned HR-LR image pairs, as represented by [4, 5, 32]. Ozcan *et al.* for the first time achieved $2.5\times$ SR of histopathological micrographs by training a CNN model with real HR-LR image pairs obtained by an automated brightfield microscope [32]. The authors of [4] and [5] used DSLR cameras to capture real-world HR and LR images at different focal lengths, and further processed them to construct real-world SR datasets called RealSR and City100, respectively. However, the methods for constructing real-world SR datasets in these works cannot be readily transferred into the application for *in situ* plankton imaging, as the underwater environment is more complex and the plankton are small and can move very fast. Moreover, the loss functions used in these works are pixel-level loss (*e.g.*, L2 loss) or content loss (*e.g.*, VGG loss), which require highly accurate image registration processing, greatly increasing the difficulty of real-underwater plankton image SR dataset construction [39].

3. Method

3.1. Real SR Image Dataset Construction

To construct the real-underwater plankton image datasets, it is necessary to simultaneously record pairs of HR and LR images of the same plankton target. In order to achieve this goal, we design a dual-channel imaging system that consists of two orthogonally-oriented telecentric imaging subsystems. Both subsystems point towards the same underwater live plankton targets in a customized quartz container through an optical beam-splitter [31]. In both subsystems, two identical digital cameras are attached to two lenses with different magnifications. They are synchronized

to capture paired snapshots of the live plankton organisms while a pulsed darkfield lighting is triggered. The captured image pairs are saved into a laptop computer immediately, and this acquisition process continues until enough numbers of paired images are obtained.

We perform an HR-LR image registration preprocessing to generate the real-underwater plankton image datasets after the raw image pairs are obtained. Here we apply several preprocessing steps described in [18] and further use template matching method to convert the raw image pairs into aligned pairs [2]. The detailed processing procedures are schematically illustrated in Figure 2 and explained as follows:

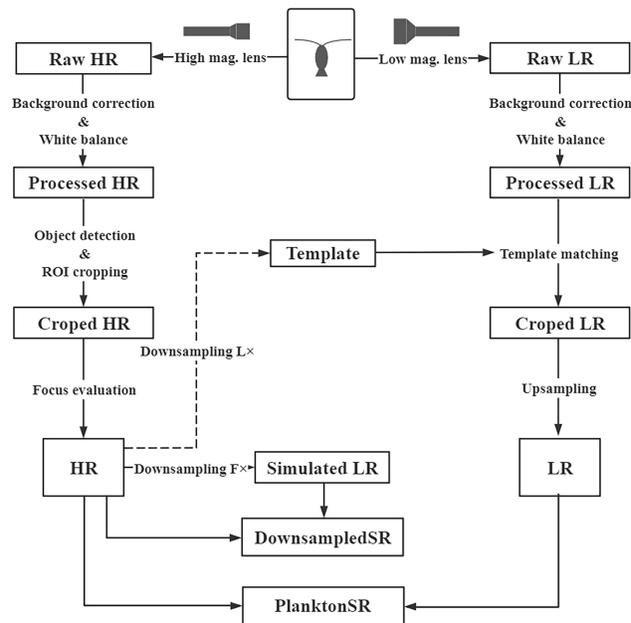


Figure 2. Procedures of the IsPlanktonSR dataset construction.

There are six post-processing steps after capturing the raw HR-LR image pairs. (1) We perform background correction to all the raw images by subtracting them with their corresponding background images. (2) We perform white-balance correction to all the raw images by the reference from a customized white target. (3) We perform object detection on the HR images and crop out the region of interest (ROIs) where the individual planktonic organisms are present. (4) We use a focus evaluation algorithm to filter the cropped ROIs and only keep the in-focus ones as the HR images in the IsPlanktonSR dataset. The first four steps are similar as in [18]. Next, (5) we downsample a clear HR ROI by a scaling factor of L (L =high lens magnification/low lens magnification) as a template to search for a matching ROI in the corresponding detected LR image, and the best-matched ROI in the raw LR image is cropped and reserved. (6) Since the magnification ratio between the raw HR-LR image pairs is not an integer, we upsample all the reserved $0.46\times$ LR images to $0.5\times$ as the final LR images in

the IsPlanktonSR dataset to satisfy the integer upsampling requirement by the pixelshuffle layer in the EDSR network [20]. Interested parties are welcome to contact us for the access to the IsPlanktonSR dataset for scientific research.

3.2. SR Network and Loss Function Selection

The SR model used in this paper is an EDSR network proposed by Lim *et al.* [20]. Its network structure is illustrated in Figure 3.

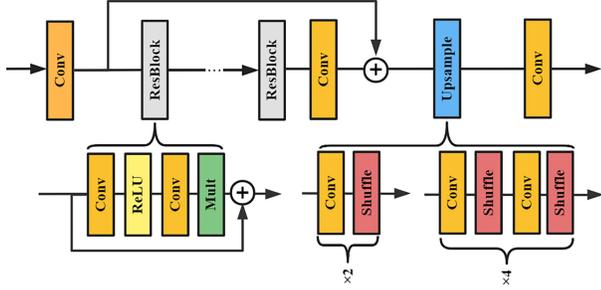


Figure 3. Network structure of the EDSR (adapted from [20]).

In the training stage of the EDSR network, we perform the L2 Loss, Perceptual Loss [14], and Contextual Loss (CX Loss) functions, and compare their SR performance. CX Loss considers an image as a collection of features, and measures the similarity between images based on the contextual similarity between features [25]. Its definition is as follows:

$$L_{CX}(\hat{y}, y, l) = \log(CX(\varphi^l(\hat{y}), \varphi^l(y))) \quad (1)$$

where φ denotes a VGG network [34], $\varphi^l(\hat{y})$ and $\varphi^l(y)$ denote the feature maps extracted from layer l of the perceptual network φ of the SR image \hat{y} and the LR image y . CX denotes the contextual similarity between the features $\varphi^l(\hat{y})$ and $\varphi^l(y)$. One of the characteristics of CX Loss is that it ignores the spatial location of features, so it allows certain imperfect alignment or local deformations between the image pairs in the training dataset.

3.3. Plankton Image SR Evaluation

We use two objective metrics of PSNR and SSIM to evaluate the SR performance of the trained model. In addition, we adopt a non-reference perceptual metric NIQE as the third indicator, as it is proven to be highly correlated with human ratings [27]. In this way, the human subjectivity can be reduced. Furthermore, we use a standard USAF1951 resolution target to quantify the resolution improvement of the trained SR model, and test its generalizability by comparing the highest resolution obtained from the LR image of the target captured by a laboratory stereoscopic microscope at $1\times$ magnification with that obtained from its SR image.

4. Experiments

4.1. IsPlanktonSR Image Dataset

By using the dual-channel imaging system with two combinations of lens magnifications ($1\times-0.5\times$) and ($2\times-0.5\times$) and the data preprocessing methods described in Section 3.1, we generate 3,453 and 5,927 registered $2\times$ and $4\times$ image pairs, which constitute the IsPlanktonSR dataset. Figure 4 presents some example image pairs of the dataset. Note all the plankton images in this paper are CLAHE enhanced [30] for better visual comfortableness. The registered image pairs are further processed as follows for the SR experiments.

$2\times$ Dataset. We keep 10 image pairs as the validation set, 20 image pairs as the test set, and the remaining pairs as the training set. Besides, we augment the training set by 4 times by horizontal flipping, vertical flipping, and horizontal and vertical flipping. After that, we extract 50% overlapping patches of 50×50 patch size and 100×100 patch size from the LR image and HR image in the training set, respectively.

It is worth mentioning that we do not keep all the extracted patches, but discard those patches containing merely dark backgrounds, because the network cannot learn any mapping for the plankton targets from them. To choose the patches containing plankton parts, we set a threshold T (default=2) in the process of patch extraction. It is only when the average pixel value of a patch is greater than this threshold can this patch be reserved. In addition, since the body size of some plankton can be larger than the DOF of the high magnification lens, some HR images are partially blurred as shown in Figure 10. It is inevitable that some patches extracted from these HR images are also blurred. We further use the focus evaluation algorithm [18] to filter out these blurred patches and only keep the sharp ones. After the above patch filtration steps, we finally obtain 84,256 HR-LR patch pairs for network training.

$4\times$ Dataset. We keep 10 image pairs as the validation set, 20 image pairs as the test set, and the remaining pairs as the training set. Similar to the process of $2\times$ dataset construction, we perform data augmentation and patch extraction on the $4\times$ dataset, and finally generate 139,313 HR-LR patch pairs, with LR patch size of 50×50 and HR patch size of 200×200 , respectively.

In addition to real dataset preparation, we downsample the HR patches in the $4\times$ dataset in IsPlanktonSR by a factor of 4 to generate a simulated dataset DownsampedSR for subsequent SR performance comparison experiment.

4.2. SR Model Training

Before the EDSR network training, we initialize it with a pre-trained model (training on the natural image dataset [35]) to speed up the convergence of network. After the tra-

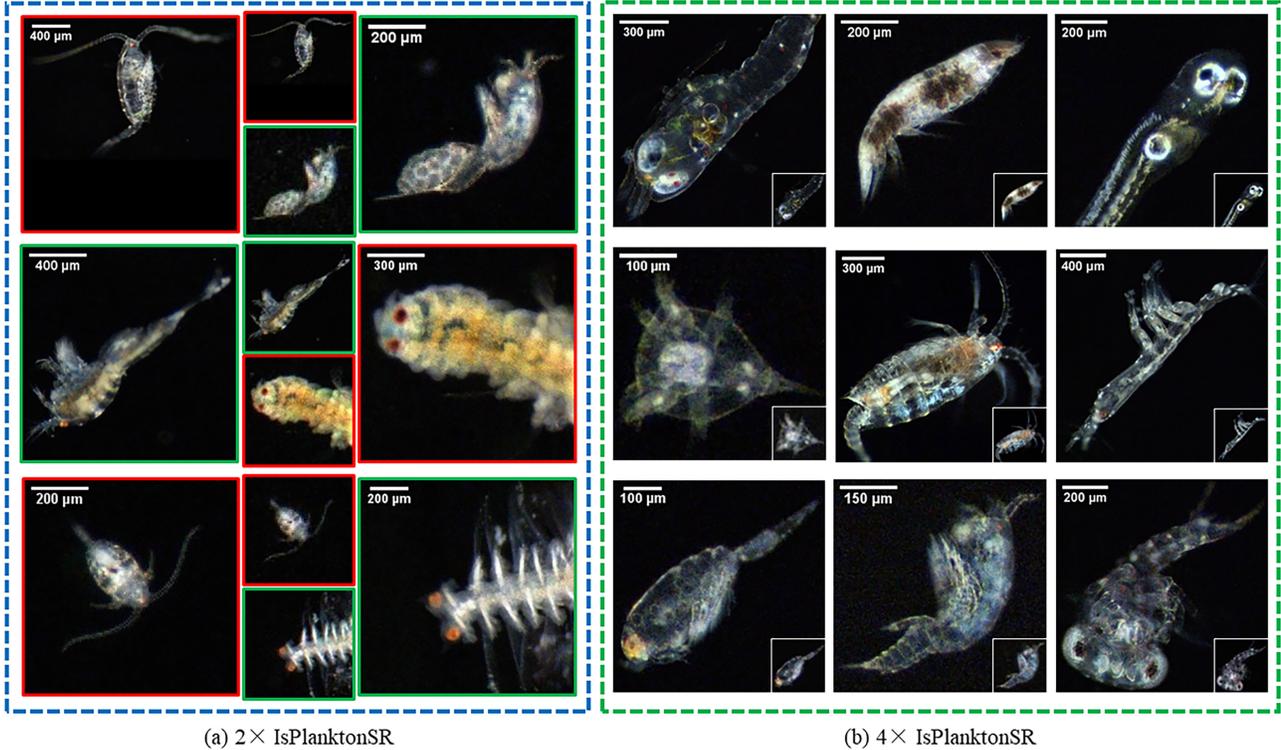


Figure 4. Example *in situ* plankton image pairs in the IsPlanktonSR dataset.

ining starts, we randomly take 16 LR patches from training set and feed them into the network at each iteration.

For training optimization, we use Adam [16] optimizer and set the initial learning rate to 4×10^{-4} and halve it at [200, 400, 500] iterations, respectively. Several different models are trained to validate the influence of adopting different loss functions and datasets on the SR performance. All experiments are conducted with Pytorch framework on a NVIDIA RTX3090 GPU server.

4.3. SR Performance Evaluation

We firstly train three $4\times$ EDSR models with different loss functions using the IsPlanktonSR dataset, and their performance evaluation results are shown in Table 1 and Figure 5. It can be seen that the images generated by the L2 Loss-trained model are smoother than the other results, although they have the highest PSNR and SSIM values. In human visual perception, the L2 Loss-trained model does not perform well as the models trained through other loss functions, and its performance on NIQE is also the worst. The output images from the Perceptual Loss-trained model show a slight improvement in terms of visual quality and NIQE compared to those from the L2 Loss-trained model, but their PSNR and SSIM values are lower than the results of the L2 Loss-trained model. In contrast, the output images of the CX Loss-trained model have the worst PSNR but the best NIQE and perception quality. Moreover, its output images recover more high-frequency details and are not as smooth

as the outputs from the other models. This observation is reasonable, as CX loss is robust to slight misalignment in the training set. Therefore, we decide to choose CX Loss to train the network.

Method	L2 Loss	Perceptual Loss	CX Loss
PSNR	33.03	32.24	30.88
SSIM	0.78	0.69	0.77
NIQE	19.91	18.47	12.61

Table 1. Quantitative results of SR on the test images from IsPlanktonSR. PSNR and SSIM (the higher, the better) are adopted for the evaluation of reconstruction accuracy; NIQE (the lower, the better) is adopted for the evaluation of visual quality.

Method	DownsampledSR	$4\times$ IsPlanktonSR
PSNR	31.01	30.88
SSIM	0.64	0.77
NIQE	17.09	12.61

Table 2. Quantitative results of SR performance on the test images from the DownsampledSR and the IsPlanktonSR datasets.

To compare the SR performance of the model trained by the simulated and the real data, we then train two $4\times$ EDSR models using the DownsampledSR and the IsPlanktonSR. As the results shown in Figure 5, the model trained by the real data recovers more high-frequency details than the model trained with the simulated data. Such visual perception is in consistent with the numerical results as shown in Table 2.

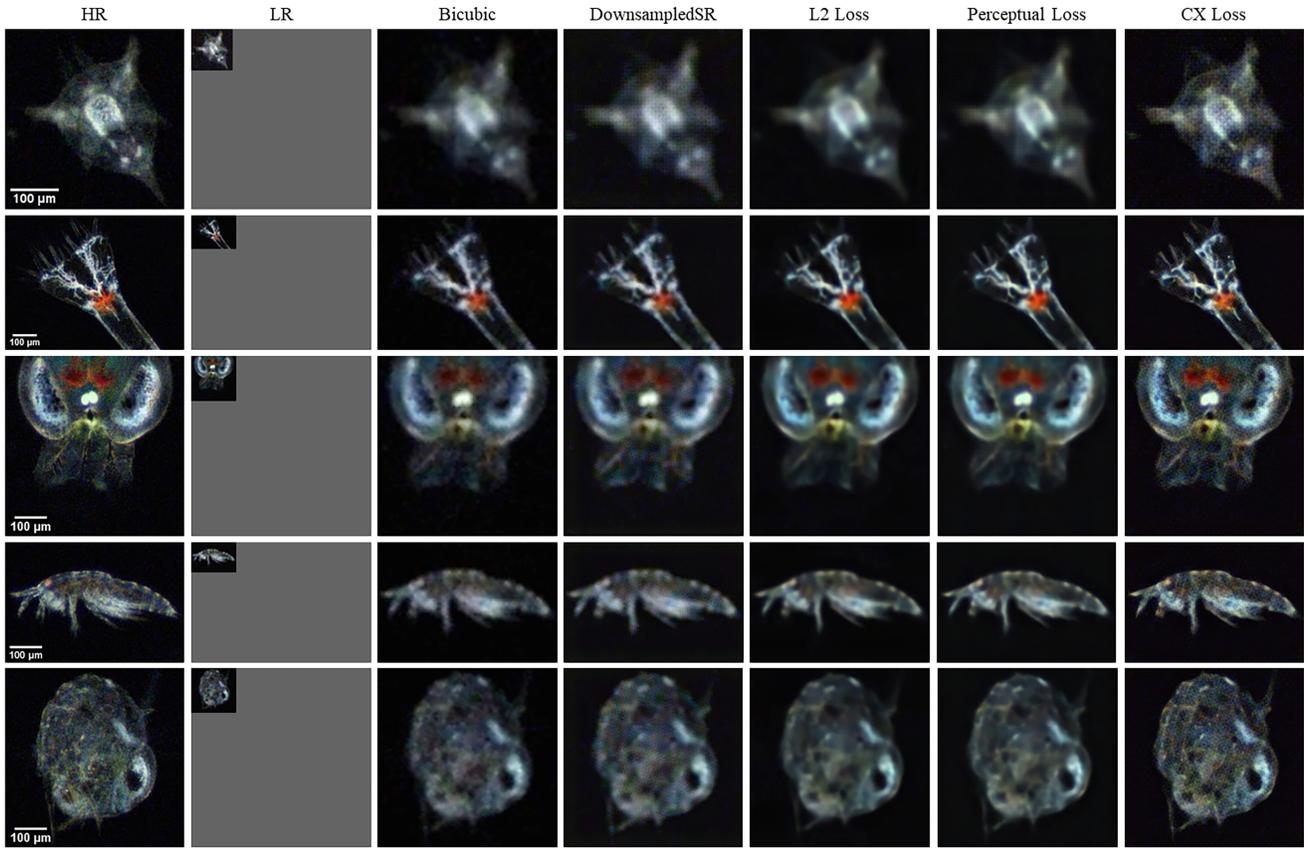
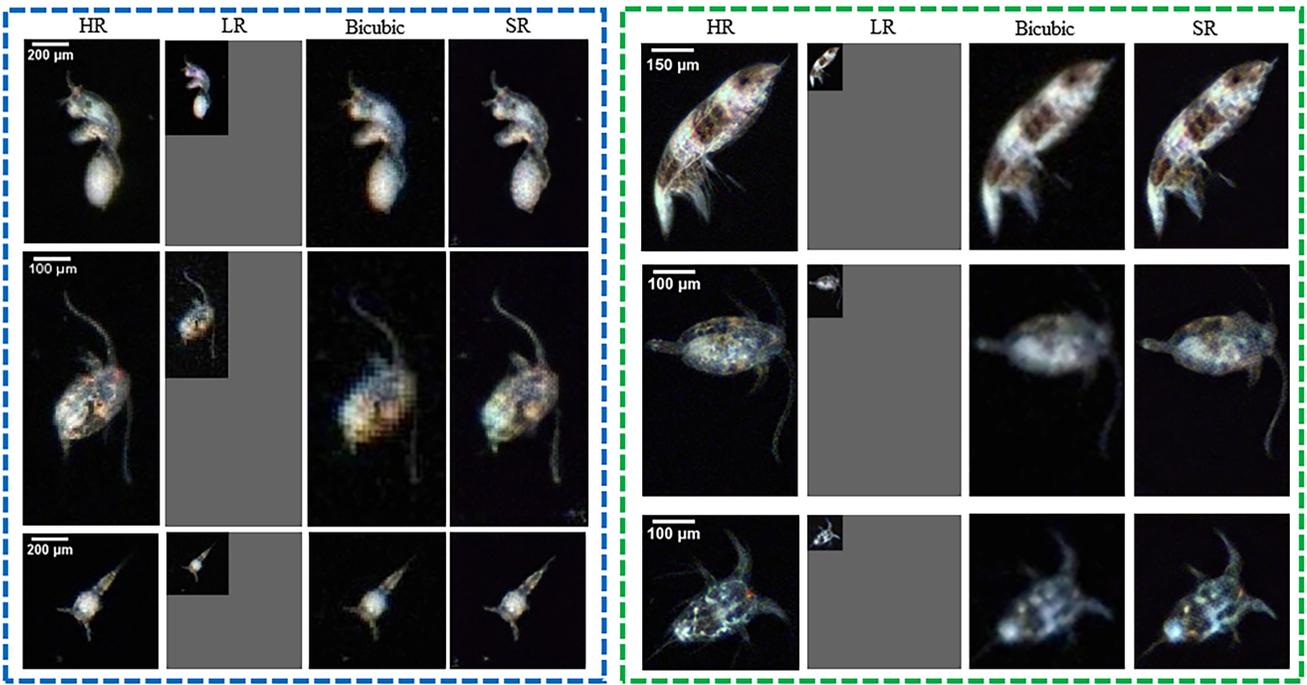


Figure 5. Visual comparison of the SR results obtained by training the $4\times$ EDSR model with different loss functions and datasets.



(a) $2\times$ IsPlanktonSR

(b) $4\times$ IsPlanktonSR

Figure 6. Visual comparison of the SR results generated by the $2\times$ and $4\times$ EDSR models trained by the IsPlanktonSR datasets.

We finally use IsPlanktonSR to train the EDSR models for $2\times$ and $4\times$ plankton SR tasks, and Figure 6 shows their SR results. It can be seen that the SR images produced by both models reveal more morphological details than their LR inputs, and have very high similarity with their HR ground truth.

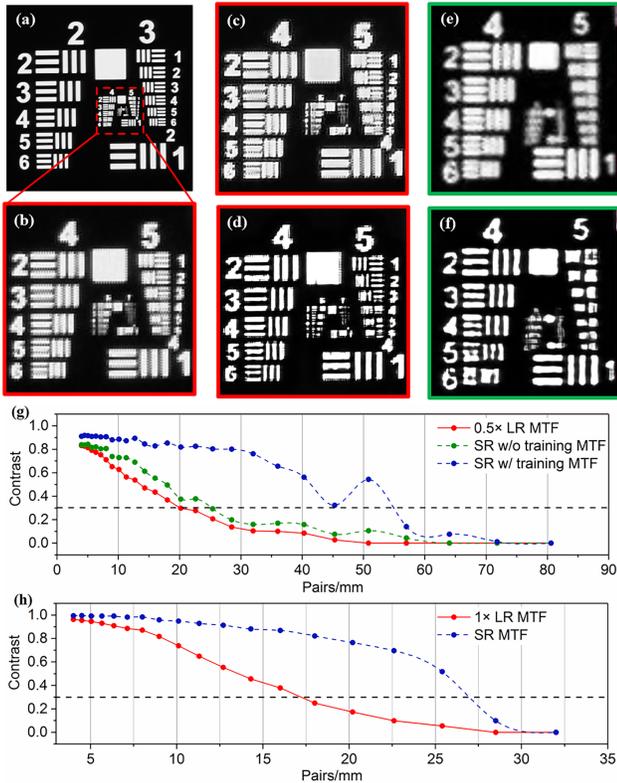


Figure 7. SR results evaluated by the USAF1951 target. (a) $0.5\times$ LR image of the target captured by the dual-channel imaging system; (b) zoom-in of the 4th and 5th elements in (a); (c) zoom-in SR image of (a) without re-training; (d) zoom-in SR image of (a) with re-training; (e) zoom-in LR image captured by a laboratory stereoscopic microscope at $1\times$ magnification; (f) SR image of (e); (g) MTFs calculated from (b), (c) and (d); (h) MTFs calculated from (e) and (f). **(Zoom in for best view)**

The results of resolution improvement evaluated by the standard USAF resolution target image are shown in Figure 7. It can be seen from the modulation transfer function (MTF) curves in Figure 7 (g) that the highest resolutions calculated from the original target LR image and its SR image are 17.96 pairs/mm and 22.63 pairs/mm, respectively. The improvement in resolution is not significant. This is because the model has not been trained to learn the mapping between the HR-LR image pairs of the target before. We then perform data augmentation and patch extraction on the target HR-LR images collected by the dual-channel imaging system, and merge the generated data into the IsPlanktonSR dataset to retrain the $4\times$ SR model. It can be seen from Figure 7 (d) that the SR image of the target generated

by the retrained model has presented much better resolution compared to its LR input. The highest resolution obtainable from the SR image has reached 50.8 pairs/mm as shown in Figure 7 (g), which is 2.82 times of that (17.96 pairs/mm) obtained in the original LR image. Moreover, the contrast of the SR images has also been significantly improved. As shown in Figure 7 (e), (f) and (h), the best resolutions obtained from the LR image of the target image taken by the stereoscopic microscope and its SR image output by the retrained model have been increased by ~ 1.59 times. The SR performance is still obvious. Note that the highest resolution obtained from the $1\times$ image is inferior to that from the $0.5\times$ image of the target. This is because the camera used on the microscope has much larger pixel size than (\sim twice) that used in the dual-imaging system.

4.4. Generalizability

In order to verify the SR effect on other unseen data, we test the generalizability of the EDSR model trained with the $4\times$ IsPlanktonSR. For the content generalization, we test some plankton images that are different from the species in IsPlanktonSR, and are collected from other sea areas. The results show that the model could still achieve good SR effect as shown in Figure 8.

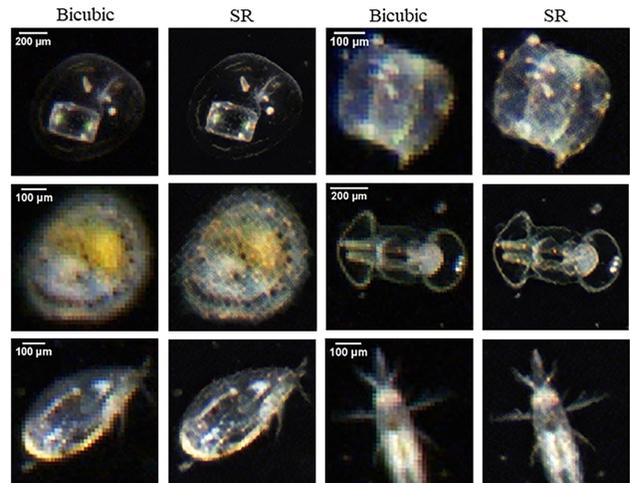


Figure 8. Visual comparison of the SR results of the EDSR model trained by the IsPlanktonSR on the test images captured by the underwater plankton imager in [18]. **(Zoom in for best view)**

For the device generalization, we conduct test on some darkfield plankton images taken by other instruments different from the dual-channel imaging system used to construct IsPlanktonSR, and the results show that the SR effects are also visually arresting even for unknown magnifications, see Figure 9. Therefore, it can be proved that the SR model trained with the IsPlanktonSR data is robust for a variety of imaging systems and plankton targets.

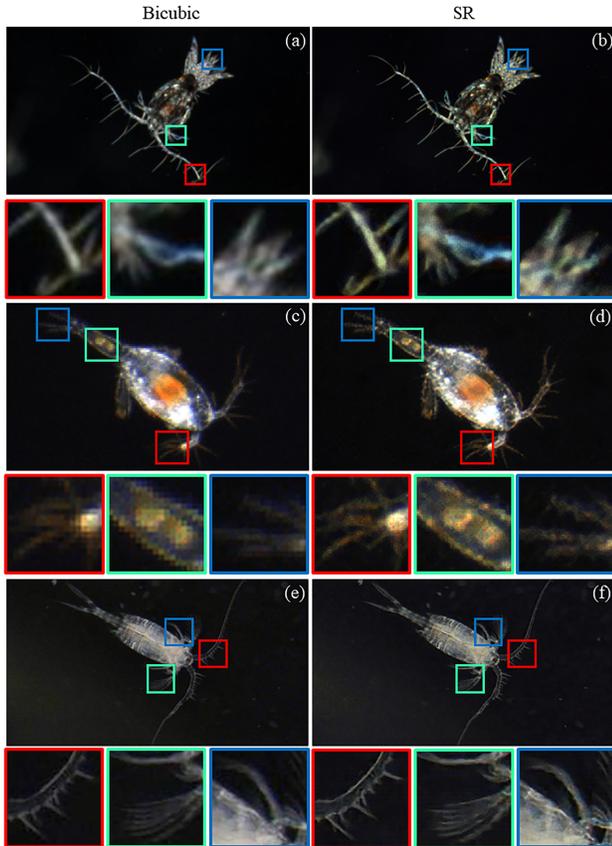


Figure 9. Visual comparison of the SR results of the EDSR model trained by the IsPlanktonSR on the test images captured by (a and b) the SPC (downloaded from <http://spc.ucsd.edu>), (c and d) the CPICS (downloaded from <https://coastaloceanvision.com>), and (e and f) a laboratory darkfield stereoscopic microscope. (**Zoom in for best view**)

4.5. Discussion

Due to the restriction of the optical imaging principle, a low magnification lens has a larger DOF than a high magnification lens. It can be seen in the IsPlanktonSR data that some plankton parts are out of the DOF in the HR images, but still in the DOF of their LR counterparts. As shown in Figure 10, the copepod in the HR image is out of focus and blurred, while its SR output even appears to be clearer than the HR ground truth. For marine plankton observation, in addition to improving the resolution, the more attractive point is that the SR image also maintains the large FOV and large DOF of the original LR images. This allows the SR images to have a larger seawater sampling volume per frame compared to the HR images with similar resolution. Taking the $4\times$ SR result from the $0.5\times$ to $2\times$ magnifications lenses achieved in this paper as an example, the sampling volume of an SR image is ~ 177 times that of its corresponding HR image. It is no doubt that the SR technology brings a significant improvement in the observational efficiency for a high magnification *in situ* imaging system.

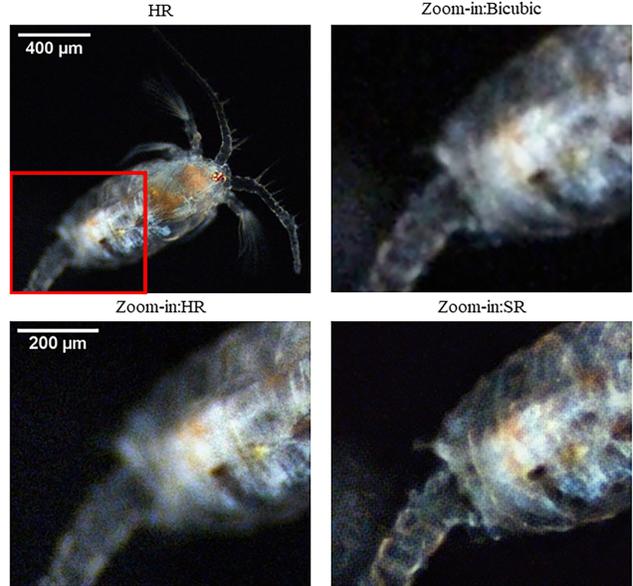


Figure 10. DOF extension by the IsPlanktonSR trained SR model.

5. Conclusion

In this paper, we construct a real-underwater plankton darkfield image dataset and use it to train a deep CNN model to achieve *in situ* plankton images SR for the first time. The method has been proved capable of alleviating the contradiction between observation area and magnification of any underwater plankton imaging system, and improving the generated plankton image resolution without sacrificing the observation volume. It is expected that the application of the SR method will help to improve the resolution of existing darkfield plankton imageries [19] captured by other instruments, inspire the future underwater plankton imager design, and eventually enhance the accuracy of plankton observations.

6. Acknowledgement

This work was supported by the Scientific Instrument Development Project of the Chinese Academy of Sciences (Grant No. YJKYYQ20190028), the Shenzhen Science and Technology Innovation Program (JCYJ20200109105823170), the National Natural Science Foundation of China (61906184), and the Shanghai Committee of Science and Technology, China (Grant No. 20DZ1100800 and 21DZ1100100).

References

- [1] Hongsheng Bi, Stuart Cook, Hao Yu, Mark C Benfield, , and Edward D Houde. Deployment of an imaging system to investigate fine-scale spatial distribution of early life stages of the ctenophore *mnemiopsis leidyi* in chesapeake bay. *Journal of Plankton Research*, 35:270–280, 2013. 1

- [2] Roberto Brunelli. Template matching techniques in computer vision: Theory and practice. 2009. 3
- [3] Dingding Cai, Ke Chen, Yanlin Qian, and Joni-Kristian Kämäräinen. Convolutional low-resolution fine-grained classification. *Pattern Recognition Letters*, 119:166–171, 2019. 2
- [4] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3086–3095, 2019. 2, 3
- [5] Chang Chen, Zhiwei Xiong, Xinmei Tian, Zheng-Jun Zha, and Feng Wu. Camera lens super-resolution. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1652–1660, 2019. 2, 3
- [6] Xuemin Cheng, Kaichang Cheng, and Hongsheng Bi. Dynamic downscaling segmentation for noisy, low-contrast in situ underwater plankton images. *IEEE Access*, 8:111012–111026, 2020. 1
- [7] Robert K. Cowen and Cedric M. Guigand. In situ ichthyoplankton imaging system (isiis): system design and preliminary results. *Limnology and Oceanography-methods*, 6:126–132, 2008. 1
- [8] Cabell S. Davis, Scott M. Gallager, and Andrew R. Solow. Microaggregations of oceanic plankton observed by towed video microscopy. *Science*, 257:230–232, 1992. 1
- [9] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2014. 3
- [10] Adam T. Greer, C. Brock Woodson, Conner E. Smith, Cedric M. Guigand, and Robert K. Cowen. Examining mesozooplankton patch structure and its implications for trophic interactions in the northern gulf of mexico. *Journal of Plankton Research*, 38:1115–1134, 2016. 1
- [11] Mary M. Grossmann, Scott M. Gallager, and Satoshi Mitarai. Continuous monitoring of near-bottom mesoplankton communities in the east china sea during a series of typhoons. *Journal of Oceanography*, 71:115–124, 2014. 1, 2
- [12] Yawen Huang, Ling Shao, and Alejandro F. Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5787–5796, 2017. 2
- [13] Jithin Saji Isaac and Ramesh Kulkarni. Super resolution techniques for medical image processing. *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, pages 1–6, 2015. 2
- [14] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European conference on computer vision*, pages 694–711. Springer, 2016. 2, 4
- [15] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016. 2
- [16] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *Computing Research Repository (CoRR)*, abs/1412.6980, 2015. 5
- [17] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, and Zehan Wang. Photo-realistic single image super-resolution using a generative adversarial network. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 105–114, 2017. 2
- [18] Jiangping Li, Tao Chen, Yangzhen Yu, Liangpei Chen, Peng Liu, Yizhou Zhang, Guangwen Yu, Jixin Chen, Haitao Li, , and Xiaohong Sun. Development of a buoy-borne underwater imaging system for in situ mesoplankton monitoring of coastal waters (accepted). *IEEE Journal of Oceanic Engineering*, 2021. 1, 2, 3, 4, 7
- [19] Jianping Li, Zhenyu Yang, and Tao Chen. Dyb-planktonnet. *IEEE Dataport*, 2021. doi: <https://dx.doi.org/10.21227/875n-f104>. 8
- [20] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017. 2, 4
- [21] Anran Liu, Yihao Liu, Jinjin Gu, Yu Qiao, and Chao Dong. Blind image super-resolution: A survey and beyond. *arXiv preprint arXiv:2107.03055*, 2021. 3
- [22] Fabien Lombard, Emmanuel Boss, Anya M. Waite, Meike Vogt, Julia Uitz, Lars Stemann, Heidi M. Sosik, Jan Schulz, Jean-Baptiste Romagnan, Marc Picheral, Jay Pearlman, Mark D. Ohman, Barbara Niehoff, Klas O. Möller, Patricia Miloslavich, Ana Lara-Lpez, Raphael Kudela, Rubens M. Lopes, Rainer Kiko, Lee Karp-Boss, Jules S. Jaffe, Morten H. Iversen, Jean-Olivier Irisson, Katja Fennel, Helena Hauss, Lionel Guidi, Gaby Gorsky, Sarah L. C. Giering, Peter Gaube, Scott Gallager, George Dubelaar, Robert K. Cowen, François Carloti, Christian Briseño-Avena, Léo Berline, Kelly Benoit-Bird, Nicholas Bax, Sonia Batten, Sakina Dorothee Ayata, Luis Felipe Artigas, and Ward Appeltans. Globally consistent quantitative observations of planktonic ecosystems. *Frontiers in Marine Science*, 6:196, 2019. 1
- [23] L. Madin, E. Horgan, S. Gallager, J. Eaton, and A. Girard. Lapis: A new imaging tool for macro-zooplankton. *OCEANS 2006*, pages 1–5, 2006. 1
- [24] Shunta Maeda. Unpaired image super-resolution using pseudo-supervision. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 288–297, 2020. 2
- [25] Roey Mechrez, Itamar Talmi, and Lihi Zelnik-Manor. The contextual loss for image transformation with non-aligned data. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 768–783, 2018. 2, 4
- [26] Ewa Merz, Thea Kozakiewicz, Marta Reyes, Christian Ebi, Peter Isles, Marco Baity-Jesi, Paul Roberts, Jules S. Jaffe, Stuart R. Dennis, Thomas Hardeman, Nelson Stevens, Tom Lorimer, and Francesco Pomati. Underwater dual-

- magnification imaging for automated lake plankton monitoring. *Water Research*, page 117524, 2021. 2
- [27] Anish Mittal, R. Soundararajan, and A. Bovik. Making a “completely blind” image quality analyzer. *IEEE Signal Processing Letters*, 20:209–212, 2013. 4
- [28] Mark D. Ohman, Russ E. Davis, Jeffrey T. Sherman, Kyle R. Grindley, Benjamin M. Whitmore, Catherine F. Nickels, and Jeffrey S. Ellen. Zooglider: An autonomous vehicle for optical and acoustic sensing of zooplankton. *Limnology and Oceanography-methods*, 17:69–86, 2019. 1
- [29] Eric C. Orenstein, Devin Ratelle, Christian Briseño-Avena, Melissa L. Carter, Peter J. S. Franks, Jules S. Jaffe, and Paul L. D. Roberts. The scripps plankton camera system: A framework and platform for in situ microscopy. *Limnology and Oceanography-methods*, 18:681–695, 2020. 1, 2
- [30] Stephen M Pizer, E Philip Amburn, John D Austin, Robert Cromartie, Ari Geselowitz, Trey Greer, Bart ter Haar Romeny, John B Zimmerman, and Karel Zuiderveld. Adaptive histogram equalization and its variations. *Graphical Models graphical Models and Image Processing computer Vision, Graphics, and Image Processing*, 39:355–368, 1987. 4
- [31] Chengchao Qu, D. Luo, E. Monari, T. Schuchert, and J. Beyrer. Capturing ground truth super-resolution data. *2016 IEEE International Conference on Image Processing (ICIP)*, pages 2812–2816, 2016. 3
- [32] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, 2017. 2, 3
- [33] Scott Samson, Lawrence Langebrake, Chad Lembke, and James Patten. Design and initial results of high-resolution shadowed image particle profiling and evaluation recorder. In *Oceans’ 99. MTS/IEEE. Riding the Crest into the 21st Century. Conference and Exhibition. Conference Proceedings (IEEE Cat. No. 99CH37008)*, volume 1, pages 58–63. IEEE, 1999. 1
- [34] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *Computing Research Repository (CoRR)*, abs/1409.1556, 2015. 4
- [35] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 114–125, 2017. 4
- [36] Nan Wang, Jia Yu, Biao Yang, Haiyong Zheng, and Bing Zheng. Vision-based in situ monitoring of plankton size spectra via a convolutional neural network. *IEEE Journal of Oceanic Engineering*, 45:511–520, 2020. 3
- [37] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018. 3
- [38] Zhihao Wang, Jian Chen, and S. Hoi. Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 2
- [39] Hao Zhang, Chunyu Fang, Xinlin Xie, Yicong Yang, W. Mei, Di Jin, and Peng Fei. High-throughput, high-resolution deep learning microscopy based on registration-free generative adversarial network. *Biomedical optics express*, 10 3:1044–1063, 2019. 2, 3
- [40] Wenlong Zhang, Yihao Liu, Chao Dong, and Y. Qiao. Ranksrgan: Generative adversarial networks with ranker for image super-resolution. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3096–3105, 2019. 3