![frontiers | Frontiers in Microbiology]

EMDS-6: Environmental Microorganism Image Dataset Sixth Version for Image Denoising, Segmentation, Feature Extraction, Classification, and Detection Method Evaluation

Peng Zhao[1], Chen Li[1]*, Md Mamunur Rahaman[1,2], Hao Xu[1], Pingli Ma[1], Hechen Yang[1], Hongzan Sun[3], Tao Jiang[4]*, Ning Xu[5] and Marcin Grzegorzek[6]

[1] Microscopic Image and Medical Image Analysis Group, College of Medicine and Biological Information Engineering, Northeastern University, Shenyang, China, [2] School of Computer Science and Engineering, University of New South Wales, Sydney, NSW, Australia, [3] Department of Radiology, Shengjing Hospital, China Medical University, Shenyang, China, [4] School of Control Engineering, Chengdu University of Information Technology, Chengdu, China, [5] School of Arts and Design, Liaoning Petrochemical University, Fushun, China, [6] Institute of Medical Informatics, University of Lübeck, Lübeck, Germany

OPEN ACCESS

Environmental microorganisms (EMs) are ubiquitous around us and have an important impact on the survival and development of human society. However, the high standards and strict requirements for the preparation of environmental microorganism (EM) data have led to the insufficient of existing related datasets, not to mention the datasets with ground truth (GT) images. This problem seriously affects the progress of related experiments. Therefore, This study develops the *Environmental Microorganism Dataset Sixth Version* (EMDS-6), which contains 21 types of EMs. Each type of EM contains 40 original and 40 GT images, in total 1680 EM images. In this study, in order to test the effectiveness of EMDS-6. We choose the classic algorithms of image processing methods such as image denoising, image segmentation and object detection. The experimental result shows that EMDS-6 can be used to evaluate the performance of image denoising, image segmentation, image feature extraction, image classification, and object detection methods. EMDS-6 is available at the https://figshare.com/articles/dataset/EMDS6/17125025/1.

Keywords: environmental microorganism, image denoising, image segmentation, feature extraction, image classification, object detection

1. INTRODUCTION

1.1. Environmental Microorganisms

Environmental Microorganisms (EMs) usually refer to tiny living that exists in nature and are invisible to the naked eye and can only be seen with the help of a microscope. Although EMs are tiny, they significantly impacts human survival (Madigan et al., 1997; Rahaman et al., 2020). Some beneficial bacteria can be used to produce fermented foods such as cheese and bread from a beneficial perspective. Meanwhile, Some beneficial EMs can degrade plastics, treat sulfur-containing waste gas in industrial, and improve the soil. From a harmful point of view,

EMs cause food spoilage, reduce crop production and are also one of the chief culprits leading to the epidemic of infectious diseases. To make better use of the advantages of environmental microorganisms and prevent their harm, a large number of scientific researchers have joined the research of EMs. The image analysis of EM is the foundation of all this.

EMs are tiny in size, usually between 0.1 and 100 microns. This poses certain difficulties for the detection and identification of EMs. Traditional "morphological methods" require researchers to look directly under a microscope (Madsen, 2008). Then, the results are presented according to the shape characteristics. This traditional method requires more labor costs and time costs. Therefore, using computer-assisted feature extraction and analysis of EM images can enable researchers to use their least professional knowledge with minimum time to make the most accurate decisions.

1.2. EM Image Processing and Analysis

Image analysis is a combination of mathematical models and image processing technology to analyze and extract certain intelligence information. Image processing refers to the use of computers to analyze images. Common image processing includes image denoising, image segmentation and feature extraction. Image noise refers to various factors in the image that hinder people from accepting its information. Image noise is generally generated during image acquisition, transmission and compression (Pitas, 2000). The aim of image denoising is to recover the original image from the noisy image (Buades et al., 2005). Image segmentation is a critical step of image processing to analyze an image. In the segmentation, we divide an image into several regions with unique properties and extract regions of interest (Kulwa et al., 2019). Feature extraction refers to obtaining important information from images such as values or vectors (Zebari et al., 2020). Moreover, these characteristics can be distinguished from other types of objects. Using these features, we can classify images. Meanwhile, the features of an image are the basis of object detection. Object detection uses algorithms to generate object candidate frames, that is, object positions. Then, classify and regress the candidate frames.

1.3. The Contribution of Environmental Microorganism Image Dataset Sixth Version (EMDS-6)

Sample collections of the EMs are usually performed outdoors. When transporting or moving samples to the laboratory for observation, drastic changes in the environment and temperature affect the quality of EM samples. At the same time, if the researcher observes EMs under a traditional optical microscope, it is very prone to subjective errors due to continuous and long-term visual processing. Therefore, the collection of environmental microorganism image datasets is challenging (Kosov et al., 2018). Most of the existing environmental microorganism image datasets are not publicly available. This has a great impact on the progress of related scientific research. For this reason, we have created the *Environmental Microorganism Image Dataset Sixth Version*

TABLE 1 | Basic information of EMDS-6 dataset, including Number of original images (NoOI), Number of GT images (NoGT).

| Class	NoOI	NoGT	Class	NoOI	NoGT
Actinophrys	40	40	Ceratium	40	40
Arcella	40	40	Stentor	40	40
Aspidisca	40	40	Siprostomum	40	40
Codosiga	40	40	K. Quadrala	40	40
Colpoda	40	40	Euglena	40	40
Epistylis	40	40	Gymnodinium	40	40
Euglypha	40	40	Gonyaulax	40	40
Paramecium	40	40	Phacus	40	40
Rotifera	40	40	Stylonychia	40	40
Vorticlla	40	40	Synchaeta	40	40
Noctiluca	40	40	-	-	-
Total	840	840	Total	840	840

(EMDS-6) and made it publicly available to assist related scientific researchers. Compared with other environmental microorganism image datasets, EMDS-6 has many advantages. The dataset contains a variety of microorganisms and provides possibilities for multi-classification of EM images. In addition, each image of EMDS-6 has a corresponding ground truth (GT) image. GT images can be used for performance evaluation of image segmentation and object detection. However, the GT image production process is extremely complicated and consumes enormous time and human resources. Therefore, many environmental microorganism image dataset does not have GT images. However, our proposed dataset has GT images. In our experiments, EMDS-6 can provide robust data support in tasks such as denoising, image segmentation, feature extraction, image classification and object detection. Therefore, the main contribution of the EMDS-6 dataset is to provide data support for image analysis and image processing related research and promote the development of EMs related experiments and research.

2. MATERIALS AND METHODS

2.1. EMDS-6 Dataset

There are 1680 images in the EMDS-6 dataset, including 21 classes of original EM images with 40 images per class, resulting in a total of 840 original images, and each original image is followed by a GT image for a total of 840. **Table 1** shows the details of the EMDS-6 dataset. **Figure 1** shows some examples of the original images and GT images in EMDS-6. EMDS-6 is freely published for non-commercial purpose at: https://figshare.com/articles/dataset/EMDS6/17125025/1.

The collection process of EMDS-6 images starts from 2012 till 2020. The following people have made a significant contribution in producing the EMDS-6 dataset: Prof. Beihai Zhou and Dr Fangshu Ma from the University of Science and Technology Beijing, China; Prof. Dr.-Ing. Chen Li and M.E. HaoXu from Northeastern University, China; Prof. Yanling Zou from Heidelberg University, Germany. The GT images of the EMDS-6
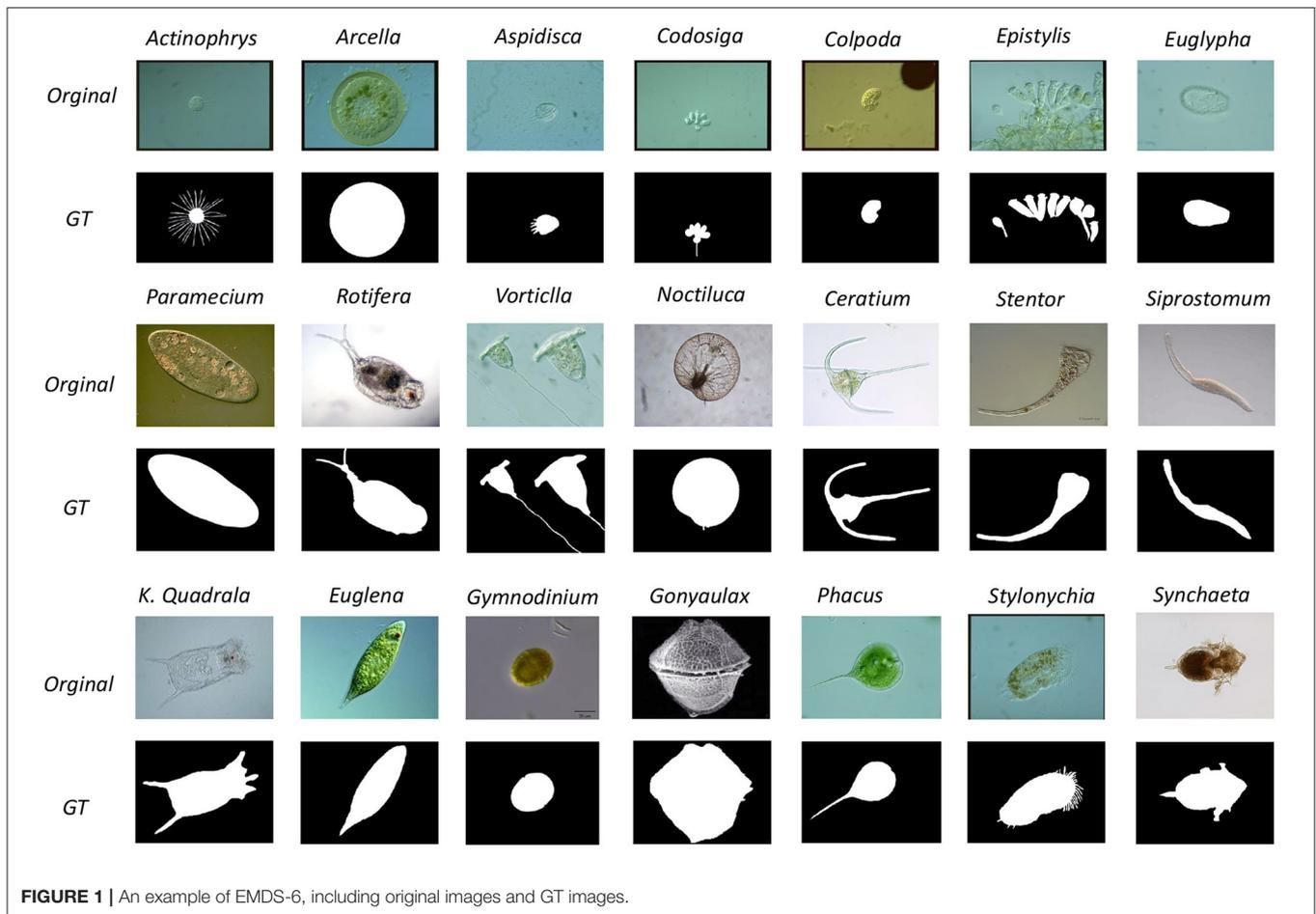
FIGURE 1 | An example of EMDS-6, including original images and GT images.

dataset are produced by Prof. Dr.-Ing Chen Li, M.E. Bolin Lu, M.E. Xuemin Zhu and B.E. Huaqian Yuan from Northeastern University, China. The GT image labeling rules are as follows: the area where the microorganism is located is marked as white as foreground, and the rest is marked as black as the background.

2.2. Experimental Method and Setup

To better demonstrate the functions of EMDS-6, we carry out noise addition and denoising experiments, image segmentation experiments, image feature extraction experiments, image classification experiments and object detection experiments. The experimental methods and data settings are shown below. Moreover, we select different critical indexes to evaluate each experimental result in this section.

2.2.1. Noise Addition and Denoising Method

In digital image processing, the quality of an image to be recognized is often affected by external conditions, such as input equipment and the environment. Noise generated by external environmental influences largely affects image processing and analysis (e.g., image edge detection, classification, and segmentation). Therefore, image denoising is the key step of image preprocessing (Zhang et al., 2022).

In this study, we have used four types of noise, Poisson noise, multiplicative noise, Gaussian noise and pretzel noise. By adjusting the mean, variance and density of different kinds of noise, a total of 13 specific noises are generated. They are multiplicative noise with a variance of 0.2 and 0.04 (marked as MN:0.2 and MN: 0.04 in the table), salt and pepper noise with a density of 0.01 and 0.03 (SPN:0.01, SPN:0.03), pepper noise (PpN), salt noise (SN), Brightness Gaussian noise (BGN), Positional Gaussian noise (PGN), Gaussian noise with a variance of 0.01 and a mean of 0 (GN 0.01–0), Gaussian noise with a variance of 0.01 and a mean of 0.5 (GN 0.01–0.5), Gaussian noise with a variance of 0.03 and a mean of 0 (GN 0.03–0), Gaussian noise with a variance of 0.03 and a mean of 0.5 (GN 0.03–0.5), and Poisson noise (PN). There are 9 kinds of filters at the same time, namely Two-Dimensional Rank Order Filter (TROF), 3 × 3 Wiener Filter [WF (3 × 3)], 5 × 5 Wiener Filter [WF (5 × 5)], 3 × 3 Window Mean Filter [MF (3 × 3)), Mean Filter with 5 × 5 Window [MF (5 × 5)]. Minimum Filtering (MinF), Maximum Filtering (MaxF), Geometric Mean Filtering (GMF), Arithmetic Mean Filtering (AMF). In the experiment, 13 kinds of noise are added to the EMDS-6 dataset image, and then 9 kinds of filters are used for filtering. The result of adding noise into the image and filtering is shown in **Figure 2**.
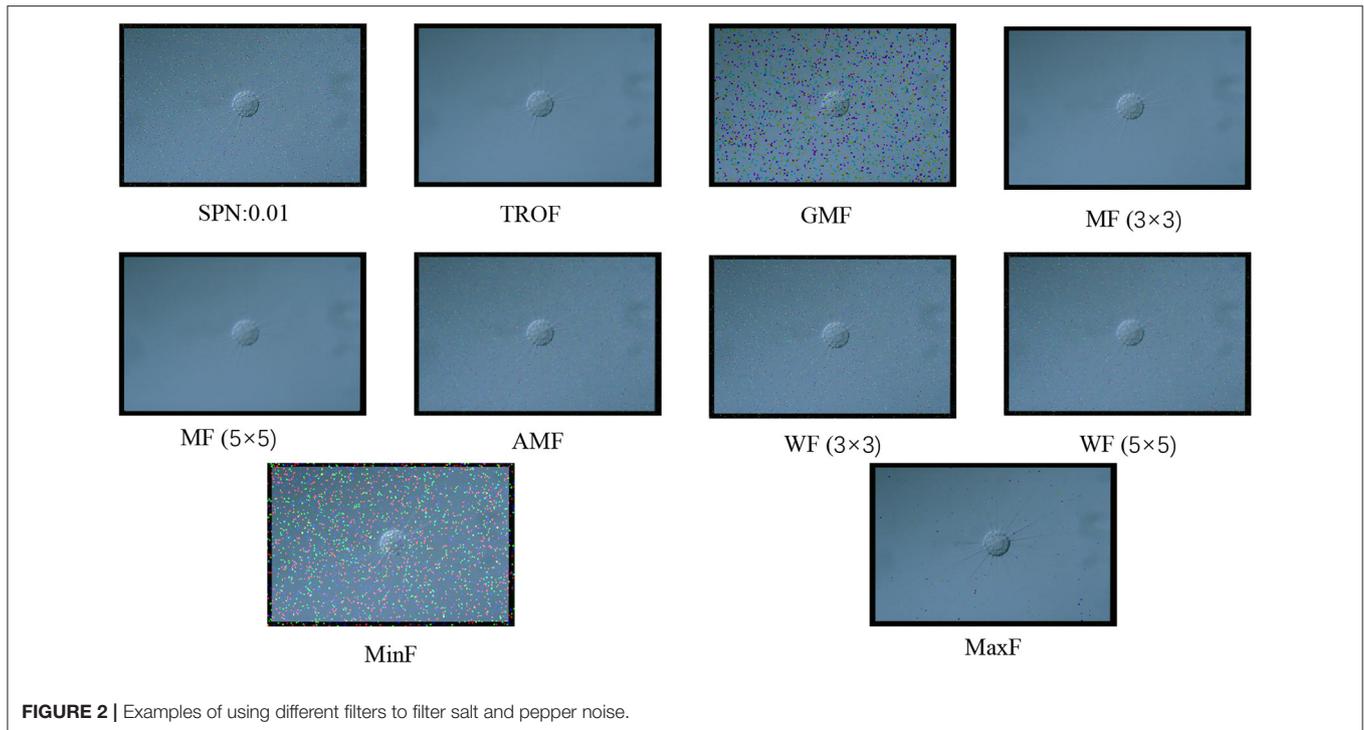
FIGURE 2 | Examples of using different filters to filter salt and pepper noise.

2.2.2. Image Segmentation Methods

This article designs the following experiment to prove that EMDS-6 can be used to test different image segmentation methods (Zhang et al., 2021). Six classic segmentation methods are used in the experiment: *k*-means (Burney and Tariq, 2014), Markov Random Field (MRF) (Kato and Zerubia, 2012), Otsu Thresholding (Otsu, 1979), Region Growing (REG) (Adams and Bischof, 1994), Region Split and Merge Algorithm (RSMA) (Chen et al., 1991) and Watershed Segmentation (Levner and Zhang, 2007) and one deep learning-based segmentation method, Recurrent Residual CNN-based U-Net (U-Net) (Alom et al., 2019) are used in this experiment. While using U-Net for segmentation, the learning rate of the network is 0.001 and the batch size is 1. In the *k*-means algorithm, the value of *k* is set to 3, the initial center is chosen randomly, and the iterations are stopped when the number of iterations exceeds the maximum number of iterations. In the MRF algorithm, the number of classifications is set to 2 and the maximum number of iterations is 60. In the Otsu algorithm, the BlockSize is set to 3, and the average value is obtained by averaging. In the region growth algorithm, we use a 8-neighborhood growth setting.

Among the seven classical segmentation methods, *k*-means is based on clustering, which is a region-based technology. Watershed algorithm is based on geomorphological analysis such as mountains and basins to implement different object segmentation algorithms. MRF is an image segmentation algorithm based on statistics. Its main features are fewer model parameters and strong spatial constraints. Otsu Thresholding is an algorithm based on global binarization, which can realize adaptive thresholds. The REG segmentation algorithm starts from a certain pixel and gradually adds neighboring pixels

according to certain criteria. When certain conditions are met, the regional growth is terminated, and object extraction is achieved. The RSMA is first to determine a split and merge criterion. When splitting to the point of no further division, the areas with similar characteristics are integrated. **Figure 3** shows a sample of the results of different segmentation methods on EMDS-6.

2.2.3. Image Feature Extraction Methods

This article uses 10 methods for feature extraction (Li et al., 2015), including two-color features, One is HSV (Hue, Saturation, and Value) feature (Junhua and Jing, 2012), and the other is RGB (Red, Green, and Blue) color histogram feature (Kavitha and Suruliandi, 2016). The three texture features include the Local Binary Pattern (LBP) (Ojala et al., 2002), the Histogram of Oriented Gradient (HOG) (Dalal and Triggs, 2005) and the Gray-level Co-occurrence Matrix (GLCM) (Qunqun et al., 2013) formed by the recurrence of pixel gray Matrix. The four geometric features (Geo) (Mingqiang et al., 2008) include perimeter, area, long-axis and short-axis and seven invariant moment features (Hu) (Hu, 1962). The perimeter, area, long-axis and short-axis features are extracted from the GT image, while the rest are extracted from the original image. Finally, we user a support vector machine (SVM) to classify the extracted features. The classifier parameters are shown in **Table 2**.

2.2.4. Image Classification Methods

In this article, we design the following two experiments to test whether the EMDS-6 dataset can compare the performance of different classifiers (Li et al., 2019; Zhao et al., 2022). Experiment 1: use traditional machine learning methods to
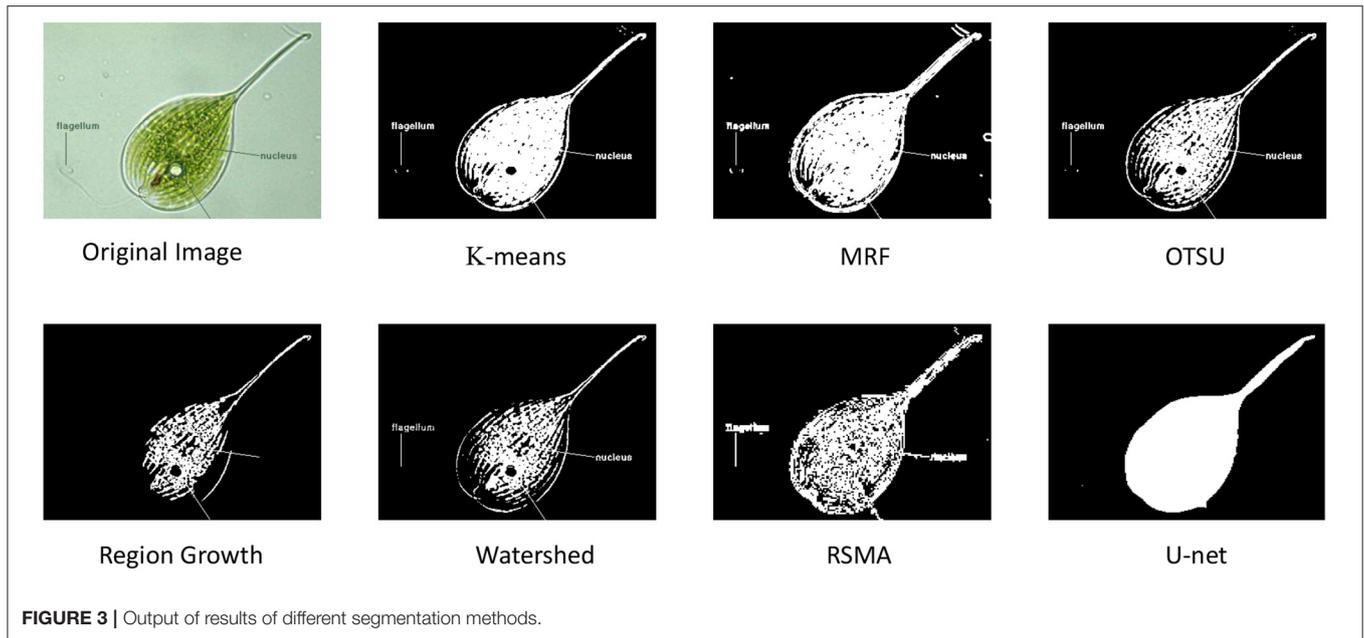
FIGURE 3 | Output of results of different segmentation methods.

TABLE 2 | Parameter setting of EMDS-6 feature classification using SVM.

Feature	Kernel	C	DFS	Tol	Max iter
LBP	rbf	50,000	ovr	1e-3	-1
GLCM	rbf	10,000	ovr	1e-3	-1
HOG	rbf	1,000	ovr	1e-3	-1
HSV	rbf	100	ovr	1e-3	-1
Geo	rbf	2,000,000	ovr	1e-3	-1
Hu	rbf	100,000	ovr	1e-3	-1
RGB	rbf	20	ovr	1e-3	-1

C, penalty coefficient; DFS, decision function shape; tol, the error value of stopping training; Geo, geometric features.

TABLE 3 | Deep learning model parameters.

Parameter	Parameter
Batch size, 32	Epoch, 100
Learning, 0.002	Optimizer, Adam

TABLE 4 | Evaluation metrics of segmentation method.

Indicators	Formula
Dice	$\frac{2 \times |V_{pred} \cap V_{gt}|}{|V_{pred}| + |V_{gt}|}$
Jaccard	$\frac{|V_{pred} \cap V_{gt}|}{|V_{pred} \cup V_{gt}|}$
Recall	$\frac{TP}{TP + FN}$

TP, True Positive; FN, False Negative; V_{pred}, the foreground predicted by the model; V_{gt}, the foreground in a GT image.

TABLE 5 | Classifier classification performance evaluation index.

Evaluation indicators	Formula
Accuracy	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	$\frac{TP}{TP+FP}$
F1-score	$2 \times \frac{P \times R}{P+R}$
Recall	$\frac{TP}{TP+FN}$

classify images. This chapter uses Geo features to verify the classifier's performance. Moreover, traditional classifiers used for testing includes, three k-Nearest Neighbor (kNN) classifiers ($k =$ 1, 5, 10) (Abeywickrama et al., 2016)], three Random Forests (RF) (tree = 10, 20, 30) (Ho, 1995) and four SVMs (kernel function = rbf, polynomial, sigmoid, linear) (Chandra and Bedi, 2021). The SVM parameters are set as follows: penalty parameter C = 1.0, the maximum number of iterations is unlimited, the size of the error value for stopping training is 0.001, and the rest of the parameters are default values.

In Experiment 2, we use deep learning-based methods to classify images. Meanwhile, 21 classifiers are used to evaluate the performance, including, ResNet-18, ResNet-34, ResNet-50, ResNet-101 (He et al., 2016), VGG-11, VGG-13, VGG-16, VGG-19 (Simonyan and Zisserman, 2014), DenseNet-121, DenseNet-169 (Huang et al., 2017), Inception-V3 (Szegedy et al., 2016), Xception (Chollet, 2017), AlexNet (Krizhevsky et al., 2012), GoogleNet (Szegedy et al., 2015), MobileNet-V2 (Sandler et al., 2018), ShuffeleNetV2 (Ma et al., 2018), Inception-ResNet - V1 (Szegedy et al., 2017), and a series of VTs, such as ViT (Dosovitskiy et al., 2020), BotNet (Srinivas et al., 2021), DeiT (Touvron et al., 2020), T2T-ViT (Yuan et al., 2021). The above models are set with uniform hyperparameters, as detailed in **Table 3**.

TABLE 6 | Similarity comparison between denoised image and original image.

ToN / DM	TROF	MF: (3 × 3)	MF: (5 × 5)	WF: (3 × 3)	WF: (5 × 5)	MaxF	MinF	GMF	AMF
PN	98.36	98.24	98.00	98.32	98.15	91.97	99.73	99.21	98.11
MN:0.2	99.02	90.29	89.45	91.98	91.08	71.15	99.02	98.89	90.65
MN:0.04	99.51	99.51	99.51	95.57	95.06	82.35	99.51	98.78	94.92
GN 0.01-0	96.79	96.45	96.13	96.75	96.40	85.01	99.44	98.93	96.28
GN 0.01-0.5	98.60	98.52	98.35	98.97	98.81	96.32	99.67	64.35	98.73
GN 0.03-0	94.64	93.99	93.56	94.71	94.71	76.46	99.05	98.74	93.82
GN 0.03-0.5	97.11	96.95	96.66	98.09	97.79	94.04	99.24	66.15	97.54
SPN:0.01	99.28	99.38	99.14	99.60	99.37	95.66	99.71	99.44	99.16
SPN:0.03	98.71	98.57	98.57	99.29	98.87	92.28	99.24	99.26	98.80
PpN	98.45	98.53	98.30	99.46	99.02	96.30	99.04	99.61	98.61
BGN	97.93	97.74	97.74	97.91	97.69	90.00	99.66	99.16	97.60
PGN	96.97	96.63	96.33	97.16	96.85	85.82	99.47	98.98	96.47
SN	97.90	97.97	97.75	99.27	98.63	99.27	98.63	99.64	98.15

ToN, types of noise; DM, denoising method. (In [%]).

TABLE 7 | Comparison of variance between denoised image and original image.

ToN / DM	TROF	MF: (3 × 3)	MF: (5 × 5)	WF: (3 × 3)	WF: (5 × 5)	MaxF	MinF	GMF	AMF
PN	1.49	0.77	1.05	0.52	0.66	3.68	2.99	0.41	0.88
MN,v: 0.2	32.49	14.94	15.65	9.33	11.36	39.22	32.49	4.32	13.35
MN,v: 0.04	10.89	10.89	10.89	2.99	3.71	14.41	10.89	0.98	4.28
GN,m: 0,v: 0.01	3.81	3.06	3.44	2.06	2.62	11.68	7.36	1.16	3.00
GN,m: 0.5,v: 0.01	0.89	0.36	0.41	0.21	0.28	0.99	1.74	61.93	0.43
GN,m: 0,v: 0.03	8.60	7.78	8.34	5.04	5.04	27.23	16.55	4.24	7.33
GN,m: 0.5,v: 0.03	1.60	1.08	1.18	0.55	0.73	2.39	3.06	56.17	1.05
SPN,d: 0.01	1.92	1.21	1.46	0.10	0.30	6.37	2.90	4.73	1.25
SPN,d: 0.03	3.84	3.39	3.39	0.33	1.09	14.64	5.18	13.02	3.15
PpN	2.88	2.18	2.44	0.17	0.72	3.72	4.48	16.84	2.09
BGN	2.35	1.63	1.94	1.09	1.38	6.67	4.57	0.84	1.66
PGN	3.79	3.04	3.42	1.67	2.13	11.56	7.33	1.23	2.98
SN	3.86	3.17	3.44	0.31	1.35	4.82	6.25	5.58	2.94

(In [%]).

2.2.5. Object Detection Method

In this article, we use Faster RCNN (Ren et al., 2015) and Mask RCNN (He et al., 2017) to test the feasibility of the EMDS-6 dataset for object detection (Li C. et al., 2021). Faster RCNN provide excellent performance in many areas of object detection. The Mask RCNN is optimized on the original framework of Faster RCNN. By using a better skeleton (ResNet combined with FPN) and the AlignPooling algorithm, Mask RCNN achieves better detection results than Faster RCNN.

In this experiment, the learning rate is 0.0001, the model Backbone is ResNet50, and the batch size is 2. In addition, we used 25% of the EMDS-6 data as training, 25% is for validation, and the rest is for testing.

2.3. Evaluation Methods

2.3.1. Evaluation Method for Image Denoising

This article uses mean-variance and similarity indicators to evaluate filter performance. The similarity evaluation index can be expressed as 1, where i represents the original image, i_1 represents the denoised image, N represents the number of pixels, and A represents the similarity between the denoised image and the original image. When the value of A is closer to 1, the similarity between the original image and the denoised image is higher, and the denoising effect is significant.

$$A = 1 - \frac{\sum_{i=1}^{n} |i_1 - i|}{N \times 255} \tag{1}$$

The variance evaluation index can be expressed as Equation (2), where S denotes the mean-variance, $L_{(i,j)}$ represents the value corresponding to the coordinates of the original image (i, j), and $B_{(i,j)}$ the value associated with the coordinates of the denoised image (i, j). When the value of S is closer to 0, the higher the similarity between the original and denoised images, the better the denoising stability.

TABLE 8 | Evaluation of feature extraction methods using EMDS-6 dataset.

Method/Index	Dice	Jaccard	Recal
k-means	47.78	31.38	32.11
MRF	56.23	44.43	69.94
Otsu	45.23	33.82	40.60
REG	29.72	21.17	26.94
RSMA	37.35	26.38	30.18
Watershed	44.21	32.44	40.75
U-Net	88.35	81.09	89.67

(In [%]).

TABLE 9 | Different results obtained by applying different features in the EMDS-6 classification experiments using SVM.

FT	LBP	GLCM	HOG
Acc	32.38	10.24	22.98
HSV	Geo	Hu	RGB
29.52	50.0	7.86	28.81

FT, Feature type; Acc, Accuracy. (In [%]).

$$S = 1 - \frac{\sum_{i=1}^{n} (L_{(i,j)} - B_{(i,j)})^2}{\sum_{i=1}^{n} L_{(i,j)}^2} \tag{2}$$

2.3.2. Evaluation Method for Image Segmentation

We use segmented images and GT images to calculate Dice, Jaccard and Recall evaluation indexes. Among the three evaluation metrics, the Dice coefficient is pixel-level, and the Dice coefficient takes a range of 0-1. The more close to 1, the better the structure of the model. The Jaccard coefficient is often used to compare the similarity between two samples. When the Jaccard coefficient is larger, the similarity between the samples is higher. The recall is a measure of coverage, mainly for the accuracy of positive sample prediction. The computational expressions of Dice, Jaccard, and Recall are shown in **Table 4**.

2.3.3. Evaluation Index of Image Feature Extraction

Image features can be used to distinguish image classes. However, the performance of features is limited by the feature extraction method. In this article, we select ten classical feature extraction methods. Meanwhile, the classification accuracy of SVM is used to evaluate the feature performance. The higher the classification accuracy of SVM, the better the feature performance.

2.3.4. Evaluation Method for Image Classification

In Experiment 1 of Section 2.2.4, we use only the accuracy index to judge the performance of traditional machine learning classifiers. The higher the number of EMs that can be correctly classified, the better the performance of this classifier. In Experiment 2, the performance of deep learning models needs to be considered in several dimensions. In order to more accurately evaluate the performance of different deep learning models, we introduce new evaluation indicators. The evaluation indexes and the calculation method of the indexes are shown in **Table 5**. In

Table 5, TP means the number of EMs classified as positive and also labeled as positive. TN means the number of EMs classified as negative and also labeled as negative. FP means the number of EMs classified as positive but labeled as negative. FN means the number of EMs classified as negative but labeled as positive.

2.3.5. Evaluation Method for Object Detection

In this article, Average Precision (AP) and Mean Average Precision (mAP) are used to evaluate the object detection results. AP is a model evaluation index widely used in object detection. The higher the AP, the fewer detection errors. AP calculation method is shown in Equations 3 and 4.

$$AP = \sum_{n=1}^{N} (r_{n+1} - r_n) Pinterp(r_{n+1}) \tag{3}$$

$$Pinterp(r_{n+1}) = max_{\hat{r} = r_{n+1}} = P(\hat{r}) \tag{4}$$

Among them, r_n represents the value of the nth recall, and $p(\hat{r})$ represents the value of precision when the recall is $\hat{r}$.

3. EXPERIMENTAL RESULTS AND ANALYSIS

3.1. Experimental Results Analysis of Image Denoising

We calculate the filtering effect of different filters for different noises. Their similarity evaluation indexes are shown in **Table 6**. From **Table 6**, it is easy to see that the GMF has a poor filtering effect for GN 0.01-0.5. The TROF and the MF have better filtering effects for MN:0.04.

In addition, the mean-variance is a common index to evaluate the stability of the denoising method. In this article, the variance of the EMDS-6 denoised EM images and the original EM images are calculated as shown in **Table 7**. As the noise density increases, the variance significantly increases among the denoised and the original images. For example, by increasing the SPN density from 0.01 to 0.03, the variance increases significantly under different filters. This indicates that the result after denoising is not very stable.

From the above experiments, EMDS-6 can test and evaluate the performance of image denoising methods well. Therefore, EMDS-6 can provide strong data support for EM image denoising research.

3.2. Experimental Result Analysis of Image Segmentation

The experimental results of the seven different image segmentation methods are shown in **Table 8**. In **Table 8**, the REG and RSMA have poor segmentation performance, and their Dice, Jaccard, and Recall indexes are much lower than other segmentation methods. However, the deep learning-based, U-Net, has provided superior performance. By comparing these image segmentation methods, it can be concluded that EMDS-6 can provide strong data support for testing and assessing image segmentation methods.

TABLE 10 | Results of experiments to classify Geo features using traditional classifiers.

Classifier type	SVM: linear	SVM: polynomial	SVM: RBF	SVM: sigmoid	RF,nT: 30
Accuracy	51.67	27.86	28.81	14.29	98.33
kNN,k: 1	kNN,k: 5	kNN,k: 10	RF,nT: 10	RF,nT: 20	–
23.1	17.86	17.38	96.19	97.86	–

(In [%]).

TABLE 11 | Classification results of different deep learning models.

Model	Precision (%)	Recall (%)	F1-score (%)	Acc (%)	PS (MB)	Time (S)
Xception	44.29	45.36	42.40	44.29	79.8	1,079
ResNet34	40.00	43.29	39.43	40.00	81.3	862
Googlenet	37.62	40.93	35.49	37.62	21.6	845
Densenet121	35.71	46.09	36.22	35.71	27.1	1,002
Densenet169	40.00	40.04	39.16	40.00	48.7	1,060
ResNet18	39.05	44.71	39.94	39.05	42.7	822
Inception-V3	35.24	37.41	34.14	35.24	83.5	973
Mobilenet-V2	33.33	38.43	33.97	33.33	8.82	848
InceptionResnetV1	35.71	38.75	35.32	35.71	30.9	878
Deit	36.19	41.36	36.23	36.19	21.1	847
ResNet50	35.71	38.58	35.80	35.71	90.1	967
ViT	32.86	37.66	32.47	32.86	31.2	788
ResNet101	35.71	38.98	35.52	35.71	162	1,101
T2T-ViT	30.48	32.22	29.57	30.48	15.5	863
ShuffleNet-V2	23.33	24.65	22.80	23.33	1.52	790
AlexNet	32.86	34.72	31.17	32.86	217	789
VGG11	30.00	31.46	29.18	30.00	491	958
BotNet	28.57	31.23	28.08	28.57	72.2	971
VGG13	5.24	1.82	1.63	5.24	492	1,023
VGG16	4.76	0.23	0.44	4.76	512	1,074
VGG19	4.76	0.23	0.44	4.76	532	1,119

Acc, Accuracy; PS, Params size.

TABLE 12 | AP and mAP based on EMDS-6 object detection of different types of EMs.

Model\sample (AP)	Actinophrys	Arcella	Aspidisca	Codosiga	Colpoda	Epistylis	Euglypha	Paramecium
Faster RCNN	0.95	0.75	0.39	0.13	0.52	0.24	0.68	0.70
Mask RCNN	0.70	0.85	0.40	0.18	0.35	0.53	0.25	0.70
Model\sample	Rotifera	Vorticella	Noctiluca	Ceratium	Stentor	Siprostomum	K.Quadrala	Euglena
Faster RCNN	0.69	0.30	0.56	0.61	0.47	0.60	0.22	0.37
Mask RCNN	0.40	0.15	0.90	0.70	0.65	0.7	0.45	0.25
Model\sample	Gymnodinium	Gonyaulax	Phacus	Stylongchia	Synchaeta	mAP	–	–
Faster RCNN	0.53	0.25	0.43	0.42	0.61	0.50	–	–
Mask RCNN	0.60	0.28	0.50	0.68	0.48	0.51	–	–

3.3. Experimental Result Analysis of Feature Extraction

In this article, we use the SVM to classify different features. The classification results are shown in **Table 9**. The Hu features performed poorly, while the Geo features performed the best. In addition, the classification accuracy of FT, LBP, GLCM, HOG, HSV and RGB features are also very different. By comparing these classification results, we can conclude that EMDS-6 can be used to evaluate image features.

3.4. Experimental Result Analysis of Image Classification

This article shows the traditional machine learning classification results in **Table 10**, and the deep learning classification results are shown in **Table 11**. In **Table 10**, the RF classifier performs the best. However, the performance of the SVM classifier using the sigmoid kernel function is relatively poor. In addition, there is a big difference in Accuracy between other classical classifiers. From the computational results, the EMDS-6 dataset is able
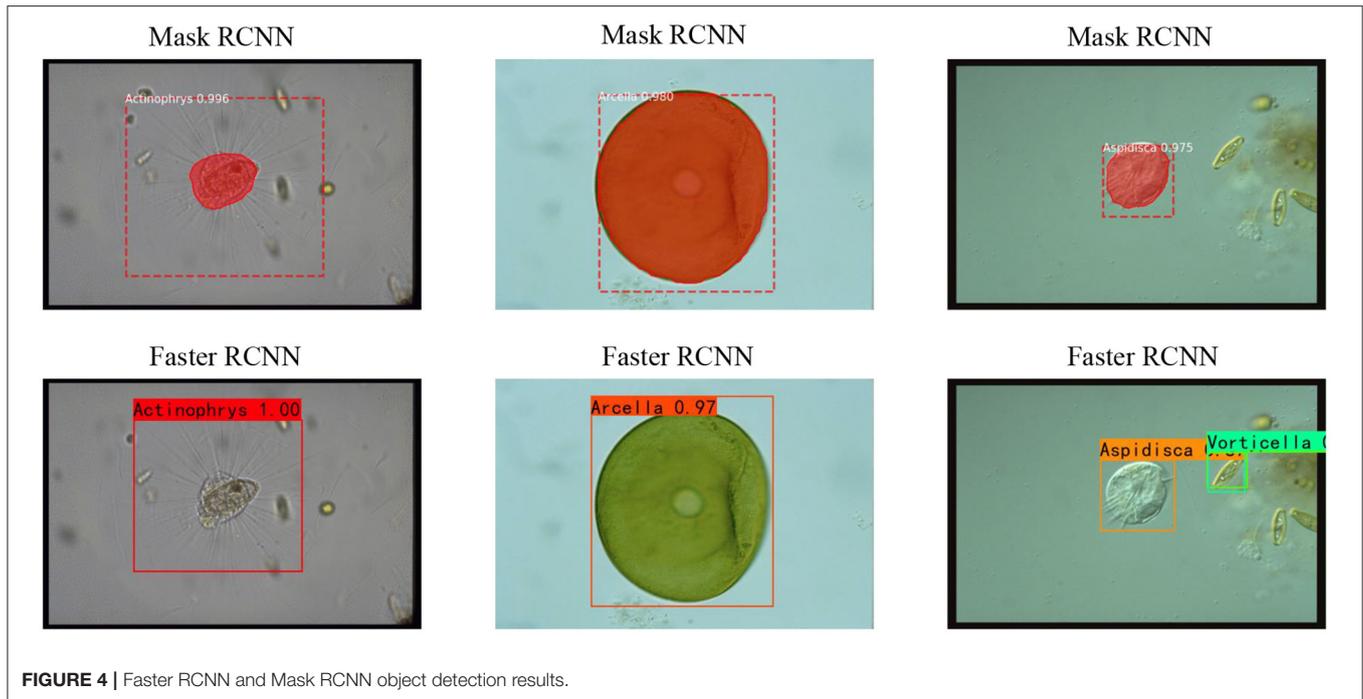
FIGURE 4 | Faster RCNN and Mask RCNN object detection results.

TABLE 13 | EMDS history versions and latest versions.

Dataset	ECN	OIN	GTIN	Dataset link	Functions
EMDS-1 (Li et al., 2013)	10	200	200	- -	IC, IS
EMDS-2 (Li et al., 2013)	10	200	200	- -	IC ,IS
EMDS-3 (Li et al., 2016)	15	300	300	- -	IC, IS
EMDS-4 (Zou et al., 2016)	21	420	420	https://research.project-10. de/em-classiffication/	IC, IS, IR
EMDS-5 (Li Z. et al., 2021)	21	420	840 (S 420, M 420)	https://github.com/ NEUZihan/EMDS-5	ID, IED, SoIS, MoIS, SoFE, MoFE, IR
EMDS-6 [In this article]	21	840	840	https://figshare.com/articles/ dataset/EMDS6/17125025/1	ID, IC, IS, IFE, IOD

IC, Image Classification; IS, Image Segmentation; SoIS, Single-object Image Segmentation; MoIS, Multi-object Image Segmentation; SoFE, Single-object Feature Extraction; MoFE, Multi-object Feature Extraction; IR, Image Retrieval; IFE, Image Feature Extraction; IOD, Image Object Detection; IED, Image Edge Detection; ID, Image denoising; ECN, EM Class Number; OIN, Original Image Number; GTIN, Ground Truth Image Number; S, Single Object; M, Multiple object.

to provide data support for classifier performance evaluation. According to **Table 11**, the classification accuracy of Xception is 44.29%, which is the highest among all models. The training of deep learning models usually consumes much time, but some models have a significant advantage in training time. Among the selected models, ViT consumes the shortest time in training samples. The training time of the ViT model is the least. The classification performance of the ShuffleNet-V2 network is average, but the number of parameters is the least. Therefore, experiments prove that EMDS-6 can be used for the performance evaluation of deep learning classifiers.

3.5. Experimental Result Analysis of Image Object Detection

The AP and mAP indicators for Faster CNN and Mast CNN are shown in **Table 12**. We can see from **Table 12** that Faster RCNN and Mask RCNN have very different object detection effects based on their AP value. Among them, the Faster RCNN model has the best effect on Actinophrys object detection. The Mask RCNN model has the best effect on Arcella object detection. Based on the mAP value, it is seen that Faster RCNN is better than Mask RCNN for object detection. The result of object detection is shown in **Figure 4**. Most of the EMs in the picture can be accurately marked. Therefore it is demonstrated that the EMDS-6 dataset can be effectively applied to image object detection.

3.6. Discussion

As shown in **Table 13**, six versions of the EMs dataset are published. In the iteration of versions, different EMSs assume different functions. Both EMDS-1 and EMDS-2 have similar functions and can perform image classification and segmentation. In addition, both EMDS-1 and EMDS-2 contain

ten classes of EMs, 20 images of each class, with GT images. Compared with the previous version, EMDS-3 does not add new functions. However, we expand five categories of EMs.

We open-source EMDSs from EMDS-4 to the latest version of EMDS-6. Compared to EMDS-3, EMDS-4 expands six additional classes of EMs and adds a new image retrieval function. In EMDS-5, 420 single object GT images and 420 multiple object GT images are prepared, respectively. Therefore EMDS-5 supports more functions as shown in **Table 13**. The dataset in this article is EMDS-6, which is the latest version in this series. EMDS-6 has a larger data volume compared to EMDS-5. EMDS-6 adds 420 original images and 420 multiple object GT images, which doubles the number of images in the dataset. With the support of more data volume, EMDS-6 can achieve more functions in a better and more stable way. For example, image classification, image segmentation, object and object detection.

4. CONCLUSION AND FUTURE WORK

This article develops an EM image dataset, namely EMDS-6. EMDS-6 contains 21 types of EMs and a total of 1680 images. Including 840 original images and 840 GT images of the same size. Each type of EMs has 40 original images and 40 GT images. In the test, 13 kinds of noises such as multiplicative noise and salt and pepper noise are used, and nine kinds of filters such as Wiener filter and geometric mean filter are used to test the denoising effect of various noises. The experimental results prove that EMDS-6 has the function of testing the filter denoising effect. In addition, this article uses 6 traditional segmentation algorithms such as k-means and MRF and one deep learning algorithm to compare the performance of the segmentation algorithm. The experimental results prove that EMDS-6 can effectively test the image segmentation effect. At the same time, in the image feature extraction and evaluation experiment, this article uses 10 features such as HSV and RGB extracted from EMDS-6. Meanwhile, the SVM classifier is used to test the features. It is found that the classification results of different features are significantly different, and EMDS-6 has the function of testing the pros and cons of features. In terms of image classification, this article designs two experiments. The first experiment uses three classic machine learning methods to test the classification performance. The second experiment uses 21 deep learning models. At the same time, indicators such as accuracy and training time are calculated to verify the performance of the model from multiple dimensions. The results show that EMDS-6 can effectively test the image classification performance. In terms of object detection, this article tests Faster RCNN and Mask RCNN, respectively. Most of the EMs in the experiment can be accurately marked. Therefore, EMDS-6 can be effectively applied to image object detection.

In the future, we will further expand the number of EM images of EMDS-6. At the same time, we will try to apply EMDS-6 to more computer vision processing fields to further promote microbial research development.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary materials, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

PZ: experiment, result analysis, and article writing. CL: data preparation, method, result analysis, article writing, proofreading, and funding support. MR and NX: proofreading. HX and HY: experiment. PM: data treatment. HS: environmental microorganism knowledge support. TJ: result analysis and funding support. MG: method and result analysis. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Abeywickrama, T., Cheema, M. A., and Taniar, D. (2016). K-nearest neighbors on road networks: a journey in experimentation and in-memory implementation. *arXiv preprint arXiv:1601.01549*. doi: 10.14778/2904121.2904125

Adams, R., and Bischof, L. (1994). Seeded region growing. *IEEE Trans Pattern Anal. Mach. Intell.* 16, 641–647.

Alom, M. Z., Yakopcic, C., Hasan, M., Taha, T. M., and Asari, V. K. (2019). Recurrent residual U-Net for medical image segmentation. *J. Med. Imaging* 6, 014006. doi: 10.1117/1.JMI.6.1.014006

Buades, A., Coll, B., and Morel, J.-M. (2005). A review of image denoising algorithms, with a new one. *Multiscale Model. Simul.* 4, 490–530. doi: 10.1137/040616024

Burney, S. M. A., and Tariq, H. (2014). K-means cluster analysis for image segmentation. *Int. J. Comput. App.* 96, 1–8.

Chandra, M. A., and Bedi, S. S. (2021). Survey on svm and their application in image classification. *Int. J. Infm. Technol.* 13, 1–11. doi: 10.1007/s41870-017-0080-1

Chen, S.-Y., Lin, W.-C., and Chen, C.-T. (1991). Split-and-merge image segmentation based on localized feature analysis and statistical tests. *CVGIP Graph. Models Image Process.* 53, 457–475.

Chollet, F. (2017). "Xception: deep learning with depthwise separable convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 1251–1258.

Dalal, N., and Triggs, B. (2005). "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)* (San Diego, CA: IEEE), 886–893.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*. Available online at: https://arxiv.53yu.com/abs/2010.11929

He, K., Gkioxari, G., Dollár, P., and Girshick, R. (2017). "Mask r-CNN," in *Proceedings of the IEEE International Conference on Computer Vision* (Honolulu, HI), 2961–2969.

He, K., Zhang, X., Ren, S., and Sun, J. (2016). "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 770–778.

Ho, T. K. (1995). "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition* (Montreal, QC: IEEE), 278–282.

Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *IRE Trans. Inform. Theory* 8, 179–187.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. (2017). "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Honolulu, HI), 4700–4708.

Junhua, C., and Jing, L. (2012). "Research on color image classification based on HSV color space," in *2012 Second International Conference on Instrumentation, Measurement, Computer, Communication and Control* (Harbin: IEEE), 944–947.

Kato, Z., and Zerubia, J. (2012). *Markov Random Fields in Image Segmentation.* Hanover, MA: NOW Publishers.

Kavitha, J., and Suruliandi, A. (2016). "Texture and color feature extraction for classification of melanoma using SVM," in *2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16)* (Kovilpatti: IEEE), 1–6.

Kosov, S., Shirahama, K., Li, C., and Grzegorzek, M. (2018). Environmental microorganism classification using conditional random fields and deep convolutional neural networks. *Pattern Recogn.* 77, 248–261. doi: 10.1016/j.patcog.2017.12.021

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Adv. Neural Inform. Process. Syst.* 25, 1097–1105.

Kulwa, F., Li, C., Zhao, X., Cai, B., Xu, N., Qi, S., et al. (2019). A state-of-the-art survey for microorganism image segmentation methods and future potential. *IEEE Access.* 7, 100243–100269.

Levner, I., and Zhang, H. (2007). Classification-driven watershed segmentation. *IEEE Trans. Image Process.* 16, 1437–1445. doi: 10.1109/TIP.2007.894239

Li, C., Ma, P., Rahaman, M. M., Yao, Y., Zhang, J., Zou, S., et al. (2021). A state-of-the-art survey of object detection techniques in microorganism image analysis: from traditional image processing and classical machine learning to current deep convolutional neural networks and potential visual transformers. *arXiv [Preprint].* arXiv: 2105.03148. Available online at: https://arxiv.org/abs/2105.03148

Li, C., Shirahama, K., and Grzegorzek, M. (2015). Application of content-based image analysis to environmental microorganism classification. *Biocybern. Biomed. Eng.* 35, 10–21. doi: 10.1016/j.bbe.2014.07.003

Li, C., Shirahama, K., and Grzegorzek, M. (2016). Environmental microbiology aided by content-based image analysis. *Pattern Anal. Appl.* 19, 531–547. doi: 10.1007/s10044-015-0498-7

Li, C., Shirahama, K., Grzegorzek, M., Ma, F., and Zhou, B. (2013). "Classification of environmental microorganisms in microscopic images using shape features and support vector machines," in *2013 IEEE International Conference on Image Processing* (Melbourne, VIC: IEEE), 2435–2439.

Li, C., Wang, K., and Xu, N. (2019). A survey for the applications of content-based microscopic image analysis in microorganism classification domains. *Artif. Intell. Rev.* 51, 577–646.

Li, Z., Li, C., Yao, Y., Zhang, J., Rahaman, M. M., Xu, H., et al. (2021). EMDS-5: Environmental microorganism image dataset fifth version for multiple image analysis tasks. *PLoS ONE* 16, e0250631. doi: 10.1371/journal.pone.0250631

Ma, N., Zhang, X., Zheng, H.-T., and Sun, J. (2018). "Shufflenet v2: practical guidelines for efficient cnn architecture design," in *Proceedings of the European Conference on Computer Vision (ECCV)* (Salt Lake City, UT), 116–131.

Madigan, M. T., Martinko, J. M., Parker, J., et al. (1997). *Brock Biology of Microorganisms,* Vol. 11. Upper Saddle River, NJ: Prentice Hall.

Madsen, E. L. (2008). *Environmental Microbiology: From Genomes to Biogeochemistry.* Oxford: Wiley-Blackwell.

Mingqiang, Y., Kidiyo, K., and Joseph, R. (2008). A survey of shape feature extraction techniques. *Pattern Recognit.* 15, 43–90. doi: 10.5772/6237

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24, 971–987. doi: 10.1109/TPAMI.2002.1017623

Otsu, N. (1979). A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybernet.* 9, 62–66.

Pitas, I. (2000). *Digital Image Processing Algorithms and Applications.* Hoboken, NJ: Wiley.

Qunqun, H., Fei, W., and Li, Y. (2013). Extraction of color image texture feature based on gray-level co-occurrence matrix. *Remote Sens. Land Resour.* 25, 26–32. doi: 10.6046/gtzyyg.2013.04.05

Rahaman, M. M., Li, C., Yao, Y., Kulwa, F., Rahman, M. A., Wang, Q., et al. (2020). Identification of covid-19 samples from chest x-ray images using deep learning: A comparison of transfer learning approaches. *J. Xray Sci. Technol.* 28, 821–839. doi: 10.3233/XST-200715

Ren, S., He, K., Girshick, R., and Sun, J. (2015). Faster r-CNN: towards real-time object detection with region proposal networks. *Adv. Neural Inform. Process. Syst.* 28, 91–99. doi: 10.1109/TPAMI.2016.2577031

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. (2018). "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Salt Lake City, UT), 4510–4520.

Simonyan, K., and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556.* Available online at: https://arxiv.53yu.com/abs/1409.1556

Srinivas, A., Lin, T.-Y., Parmar, N., Shlens, J., Abbeel, P., and Vaswani, A. (2021). Bottleneck transformers for visual recognition. *arXiv preprint arXiv:2101.11605.* Available online at: https://arxiv.org/abs/2101.11605

Szegedy, C., Ioffe, S., Vanhoucke, V., and Alemi, A. (2017). "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI Conference on Artificial Intelligence* (San Francisco, CA).

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., et al. (2015). "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Boston, MA), 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (Las Vegas, NV), 2818–2826.

Touvron, H., Cord, M., Douze, M., Massa, F., Sablayrolles, A., and Jégou, H. (2020). Training data-efficient image transformers & distillation through attention. *arXiv preprint arXiv:2012.12877.* Available online at: https://arxiv.org/abs/2012.12877

Yuan, L., Chen, Y., Wang, T., Yu, W., Shi, Y., Tay, F. E., et al. (2021). Tokens-to-token vit: training vision transformers from scratch on imagenet. *arXiv preprint arXiv:2101.11986.* Available online at: https://arxiv.53yu.com/abs/2101.11986

Zebari, R., Abdulazeez, A., Zeebaree, D., Zebari, D., and Saeed, J. (2020). A comprehensive review of dimensionality reduction techniques for feature selection and feature extraction. *J. Appl. Sci. Technol. Trends* 1, 56–70. doi: 10.38094/jastt1224

Zhang, J., Li, C., Kosov, S., Grzegorzek, M., Shirahama, K., Jiang, T., et al. (2021). Lcunet: A novel low-cost u-net for environmental microorganism image segmentation. *Pattern Recognit.* 115, 107885. doi: 10.1016/j.patcog.2021.107885

Zhang, J., Li, C., Rahaman, M., Yao, Y., Ma, P., Zhang, J., et al. (2022). A comprehensive review of image analysis methods for microorganism counting: from classical image processing to deep learning approache. *Artif. Intell. Rev.* 55, 2875–2944. doi: 10.1007/s10462-021-10082-4

Zhao, P., Li, C., Rahaman, M., Xu, H., Yang, H., Sun, H., et al. (2022). A comparative study of deep learning classification methods on a small environmental microorganism image dataset (emds-6): From convolutional neural networks to visual transformers. *arXiv [Preprint].* arXiv: 2107.07699. Available online at: https://arxiv.org/pdf/2107.07699.pdf

Zou, Y. L., Li, C., Boukhers, Z., Shirahama, K., Jiang, T., and Grzegorzek, M. (2016). "Environmental microbiological content-based image retrieval system using

internal structure histogram," in *Proceedings of the 9th International Conference on Computer Recognition Systems*, 543–552.