An Enhanced Framework of Generative Adversarial Networks (EF-GANs) for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

HAO XU[1,2], CHEN LI[1,2], MD MAMUNUR RAHAMAN[1,2], YUDONG YAO[3], (Fellow, IEEE), ZIHAN LI[1,2], JINGHUA ZHANG[1,2], FRANK KULWA[1,2], XIN ZHAO[4], SHOULIANG QI[1,2], (Member, IEEE), AND YUEYANG TENG[1,2]

[1]Microscopic Image and Medical Image Analysis Group, MBIE College, Northeastern University, Shenyang 110819, China
[2]Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education, Shenyang 110819, China
[3]Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA
[4]Environmental Engineering Institute, Northeastern University, Shengyang 110189, China

Corresponding author: Chen Li (lichen201096@hotmail.com)

ABSTRACT The main obstacle to image augmentation with *Generative Adversarial Networks* (GANs) is the need for a large amount of training data, but this is difficult for small datasets like *Environmental Microorganisms* (EMs). EM image analysis plays a vital role in environmental monitoring and protection, but it is often encountered with small datasets due to the difficulty of EM image collection. To this end, we propose an *Enhanced Framework of GANs* (EF-GANs) that combines geometric transformation methods and GANs for EM image augmentation. First of all, the color of an EM image has an insignificant impact on its class label, based on this fact, we perform color space augmentation to the original EM images. Secondly, we train EF-GANs with augmented EM images to generate utterly new EM images. Finally, we rotate the generated samples in various directions to obtain a more natural performance. In this study, we use VGG16 and ResNet50 networks to evaluate the proposed EF-GANs on 21 different types of EMs (420 EM images). It is observed that the average precision (AP) of VGG16 increases between 4.5% and 84.1% in 20 EM classes and one class remains unchanged. The AP of Resnet50 rises between 8.7% and 38.7% in 12 EM classes and reaches 100% in two EM classes. Furthermore, to reflect the generalization performance of EF-GANs, we employ an utterly new EM image dataset (630 EM images) to test the previous VGG16 networks. We select the VGG16 networks with original and optimal settings for all the EM classes, and for testing, optimal settings for a single EM class is considered. In the 20 of 21 one-vs-rest EM image classification tasks, the AP of VGG16 increases between 1.66% and 88.1%. The results demonstrate that the proposed EF-GANs can achieve outstanding performance in augmenting single EM images with high quality and resolution, thus, to improve the APs of EM image classification.

INDEX TERMS Image augmentation, environmental microorganism, microscopic image, small dataset, generative adversarial networks, image analysis, image classification.

I. INTRODUCTION

In recent years, with the continuous progress of industry, there have been numerous environmental problems, such as water

The associate editor coordinating the review of this manuscript and approving it for publication was Kumaradevan Punithakumar.

pollution and what is the full form of particulate matter (PM) cases, increasing the risk of diseases. Instead of using chemicals to eliminate such pollutants, a more harmless approach would be taking advantage of the natural consumption of *Environmental Microorganisms* (EMs). EMs are microscopic living organisms in natural and artificial environments

IEEE *Access*

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

(e.g., forests and farmlands), which are useful for cleaning environments [1]. For example, *Actinophrys* can digest organic waste in sludge and increase the quality of freshwater, whereas *Rotifera* can decompose rubbish in water and reduce the level of eutrophication. To have more knowledge of the EMs, microorganism classification is a primary and significant task in microbiological fields. Generally, image analysis techniques are widely used in EM classification tasks because they are rapid, low cost, and objective process [2] [3] [4].

Since EMs are tiny and invisible to the naked eyes (their size usually varies between 0.1 and 100 μm), they can only be observed under microscopes. EM images are captured using a digital camera embedded in the microscope (p.s., no matter how a microorganism is in any direction in the image, it is still a microorganism. Therefore, the microorganism is rotation-invariant.). Since most of the EMs are colorless and transparent, it is usually challenging to see EMs' outlines and details under the microscope. Therefore, it is necessary to dye the microorganisms with an additional light source so that we can make the outlines and details of the microorganisms clear, and the additional light source does not affect the class label of microorganisms. However, there are some difficulties in collecting EM images. EMs are different from the laboratory microorganisms (LMs). LMs grow in the medium and there is usually only one type of microorganism in one medium. Therefore, the background of the image is pure and free of impurities. On the contrary, EMs grow in natural environments (e.g., forests, rivers, fish ponds, and farmlands). When the soil, water, and other samples collected in the natural environment are observed under the microscope, there are inevitably different types of microorganisms in the sample, so finding one microorganism of a specific type in a visual field is difficult. So, collecting EM images is difficult and the datasets are normally small. Therefore, although image analysis based on EM classification methods are effective, they often suffer from a small training dataset problem [5]. Especially, novel deep learning methods are sensitive to this problem even more severe than classical machine learning methods (e.g., support vector machine (SVM)). When the training dataset is unbalanced (the amount of available data is not equal between different categories), image classification accuracy often decreases significantly [6]. To this end, data augmentation techniques are normally used to expand the dataset to obtain a better training performance.

Traditionally, data augmentation methods include image cropping, rotation, flipping, scaling, translation, contrast transformation, color space transformation, and noise injection [7]. In contrast, the novel Generative Adversarial Networks (GANs) can generate more natural and vivid images [7]. Hence, more and more works select GANs for data augmentation. For example, in [8], CycleGAN is used in an emotion classification task for human face image generation. However, for small datasets such as EM microscopic images, directly using GANs to perform image augmentation has the following two problems: First, due to the various directional properties of EM microscopic images, it is

difficult to generate various directional microscopic images by using a few various directional EM microscopic images. Only with a large number of images covering sufficient directions can GANs generate microscopic images in various directions. Second, because of the small dataset problem, if we directly use GANs to achieve image augmentation, many details will be missed from generated images. Without a large number of images with enough details, GANs cannot generate images with sufficient details. For these two problems, we propose an *Enhanced Framework of GANs* (EF-GANs) to solve its various directional small dataset problem. Our model is not a new GAN, but an enhanced framework of GANs. It means that all the GANs can be inserted into our model.

The difference between EF-GANs and GANs is that in addition to discriminator and generator networks that make up GANs, EF-GANs adds two heuristic steps at the beginning and the end of GANs. Before inputting the EM images into GANs, in step-1, we rotate EM images to unify the direction of the same class of EMs in the images and then perform color space transformation to these images. Because EM images are rotation-invariant, unifying the direction of EMs is able to solve the problem that GANs are difficult to generate EM images of different directions without enough images covering a number of directions. So, GANs only need to generate EM images in one direction. The operation of color space transformation makes the number of images input to GANs meet the training needs of GANs, which can solve the problem of insufficient training images of GANs. It enables GANs to generate more detailed images. Finally, after GANs generate the images, in step-3, we rotate these images in various directions. Because we rotate EMs images to a unified direction in step-1 and lose the direction information of EMs, we finally need to make direction compensation. So, we rotate the image in many different directions and obtain many generated EM images. Hence, it makes up for the loss of EM directional features after the augmentation of GANs.

The workflow of EF-GANs is shown in Fig. 1 and the process of EF-GANs is as follows:

- In step-1, to train GANs effectively, we initially augment each image 15 times. In (I), original EM images and their Ground Truth (GT) images are prepared. In (II), we localize the positions of EMs in the GT images to find the directions of their main axis and minimum bounding rectangles. In (III) and (IV), we use the MBRs to crop the EMs in the original images and rotate the images to make EMs' main axis horizontal. In (V), because the colors of images have little effect on their class label [9] [10], we perform color space augmentations on these images.
- In step-2, we input the images from step-1 (V) into the GANs to generate images, which augment EM images four times.
- In step-3, we rotate the results from step-2 into various directions to generate more natural EM images, which augment EM images four times.
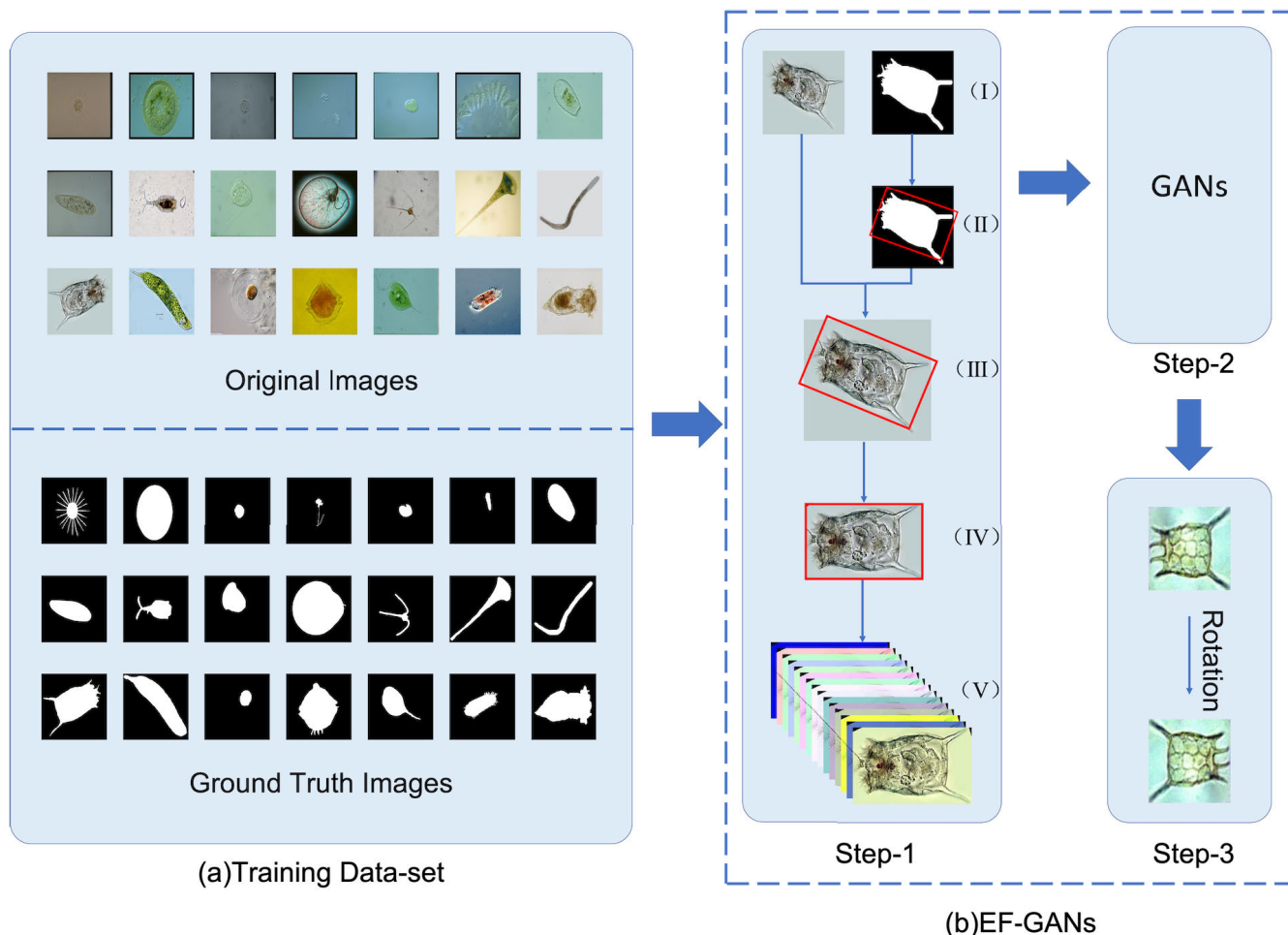
H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

IEEE *Access*

FIGURE 1. The workflow of the proposed EF-GANs for EM image augmentation. (a) is the training dataset. (b) is the framework of EF-GANs. Step-1, step-2, and step-3 are three steps of EM image augmentation. In step-1, first of all, we use the GT images to acquire the long axis (main axis), the short axis, the centroid, and the minimum bounding rectangle (MBR) of their corresponding original images. Secondly, we cut out the outer part of the MBR of the original images. And then, we rotate the cropped images to make the main axis horizontal for unifying the position of the same class of EMs. Finally, we perform color space transformation to the rotated images. In step-2, we input the images of step-1 to GANs to generate more images. In step-3, to simulate EM images in the real environments, we rotate generated images in various directions.

The main contribution of this paper is as follows:

- We utilize the location information of GT images and transform each of EM images in the same direction (main axis direction) to unify the direction of EMs, so that GANs can generate EM images in the same direction.
- Based on the fact that the backgrounds of EM images have different colors under different light sources. Therefore we perform color space transformation to the EM images to reproduce them under different color light sources in the real environments. This technique expands the images and does not affect the quality and class label of images.
- We combine the traditional image processing technology with GANs to build the EF-GANs model, which can generate high-quality EM images with fewer images. Therefore, we have resolved the requirement of a large amount of training data.
- The generated images are in any directions and can improve the classification APs of deep learning networks.

The structure of this paper is as follows: Sec. II is the related work about existing image augmentation methods. Sec. III gives a detailed description of EF-GANs. Sec. IV introduces the experiment settings and evaluation of augmented images. Sec. V closes this paper with a brief conclusion and introduces our future work.

II. RELATED WORK

In this section, image augmentation approaches are introduced, including basic image manipulation and deep learning based methods.

A. IMAGE MANIPULATION BASED IMAGE AUGMENTATION

1) GEOMETRIC TRANSFORMATIONS

Geometric transformations are traditional data augmentation methods. The effectiveness of geometric transformation relates to their safety, referring to the likelihood of preserving the label post-transformation. For example, rotations and flips are generally safe on ImageNet [11] challenges such as "cat vs dog," but not safe for digit recognition tasks such as

IEEE *Access*

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

"6 vs 9." Geometric transformations include rotation, flipping, cropping, and translation [7].

- Rotation augmentations are done by rotating the image clockwise or counterclockwise between 1° and 359°. The safety of rotation augmentations is heavily determined by the rotation degree parameter. Slight rotations such as between 1° and 20° could be useful on digit recognition tasks such as MNIST [12], but as the rotation degree increases, the safety of rotation augmentations would be destroyed.
- Flipping augmentations include horizontal and vertical axis flipping. Generally, horizontal axis flipping is much more common than vertical flipping. For example, images of different kinds of animals like cat and dog can generally perform data augmentation by horizontal axis flipping and it does not break the label of images.
- Cropping augmentations refer to cutting out the edges of an image. Its destination is to remove some of the background information and retain the main information of the image. Cropping augmentations are generally safe. It does not break the label of images.
- Translation augmentations are done by shifting images up, down, left, or right. As the original image is translated in a direction, the remaining space can be filled with either a constant value such as 0 s or 255 s, or it can be filled with random or Gaussian noise. This operation does not break the label of images. In addition, translation augmentations are similar to cropping augmentations. They both retain the main information of images.

Thus, geometric transformations are reasonable solutions to fix the biases that are presented in the training data, and this can be implemented quickly. However, geometric transformations have some disadvantages, such as the requirement of additional memory, transformation computational costs, and additional training time. Moreover, some geometric transformations like rotation and flipping may not keep the label post-transformations. Therefore, the scope of where and when geometric transformations can be applied is relatively limited.

2) COLOR SPACE TRANSFORMATIONS

An image is composed of RGB color space metrics. Generally, color space transformations are done by each color space metric by multiplying a random number. However, color transformations may lose important color information and they are not always preserving the label post-augmentations [13]. For example, in [14], the color of blood is the most important feature of distinguishing blood from water or plant. Therefore, color transformations would result in that the model cannot recognize the blood in the image.

3) KERNEL FILTER

Kernel filters are used to sharpen and blur images [9] [10]. In [15], they use a unique kernel filter that randomly swaps the pixel values in a sliding window. They call this augmentation technique PatchShuffle Regularization. They show that PatchShuffle improves the generalization ability of convolutional neural networks (CNN), especially for small datasets.

4) MIXING IMAGES

Mixing images is a very counterintuitive approach to data augmentation. After the image transformation, the image cannot be recognized. In [16], they propose a SamplePairing method. Firstly, they perform primary image augmentations such as flipping for two images which are randomly selected from training datasets, and then take the average of pixels of two images as a new sample. The two images are not even limited to the same category. SamplePairing is very simple and useful for medical images. SamplePairing significantly improved classification accuracy for all the test sets. For example, the top-1 error rate was reduced from 8.22% to 6.93% on CIFAR-10 [17].

B. DEEP LEARNING BASED IMAGE AUGMENTATION
1) FEATURE SPACE AUGMENTATION

Neural networks are very powerful at mapping high dimensional inputs into lower dimensional representations. In [18], they increase the classification accuracy on CIFAR-100 from 66% to 73% accuracy by manipulating the modularity of neural networks to isolate and refine individual layers after training. Synthetic Minority Over-sampling Technique (SMOTE) [19] is a popular augmentation used to alleviate problems with class imbalance.

2) ADVERSARIAL ATTACK

Adversarial attack is a popular technique in image processing. Adversarial attack refers to adding adversarial noise to the original image to make the model misclassify. In [20], they use DeepFool to achieve adversarial attack and cause a misclassification with high confidence. In [21], they can misclassify 70.97% of images by changing one pixel. In [22], they cause misclassifications with adversarial attack limited to the border of images. The success of adversarial attack is especially exaggerated as the resolution of images increases.

3) GAN-BASED DATA AUGMENTATION

The use of GANs to perform data augmentation in medical imaging shows good results. In [23], a cervical intraepithelial neoplasia grade classification problem is investigated on segmented epithelium image patches. A conditional GAN is applied to expand the limited training dataset by synthesizing realistic cervical histopathology images. To control the feature quality of selected synthetic images for data augmentation, they propose a synthetic-image filtering mechanism based on the divergence in feature space between generated images and class centroids. As a result, they increase the classification accuracy from 66.3% to 71.7% using the same ResNet18 baseline classifier after leveraging conditional GAN generated images with feature based filtering.

In [24], a method is proposed for generating liver lesion images using deep learning GAN. This method is

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

IEEE *Access*

demonstrated on a limited dataset of computed tomography (CT) images of 182 liver lesions (Cysts, Metastases, and Hemangiomas). It achieves an improvement of 7% in accuracy for the liver lesion classification task by using synthetic data augmentation, showing that the generated medical images can work for synthetic data augmentation, and improve the performance of CNN for medical image classification. In [25], a generative algorithm is proposed to produce synthetic abnormal brain tumor multi-parametric MRI images from their corresponding segmentation masks using an image-to-image translation GAN. The dice score (mean/standard deviation) with the experimental setting of "Real+Synthetic" is 0.82/0.08. That is a good result. In [26], a novel approach is introduced to generate synthetic medical images using generative adversarial networks (GANs). The proposed model can create brain PET images for three different stages of Alzheimer's disease—normal control (NC), mild cognitive impairment (MCI), and Alzheimer's disease (AD). The mean SSIM of this model of real and generated images is 77.48, showing good generating performance.

In [27], Conditional Progressive Growing of Generative Adversarial Networks (CPGGANs) is proposed for medical image data augmentation using automatic bounding box annotation to improve the training robustness, incorporating highly-rough bounding box conditions incrementally into Progressive Growing of Generative Adversarial Networks (PGGANs) to place brain metastases at desired positions/sizes on 256×256 Magnetic Resonance (MR) images, for Convolutional Neural Network-based tumor detection. The novelty of this study is very good. The work of [28] proposes a two-step GAN-based data augmentation method that generates and refines brain Magnetic Resonance (MR) images with/without tumors separately. In the first step, PGGANs are used to generate realistic/diverse 256×256 images. In the second step, the Multimodal UNsupervised Image-to-image Translation (MUNIT) is used to combine GANs/Variational AutoEncoders or SimGAN, to further refine the texture/shape of the PGGAN-generated images to the real ones. The results show that the sensitivity of this study is between 93.67% and 97.48% in tumor detection. The work of [29] evaluates the use of CycleGAN for data augmentation in CT segmentation tasks, where a CycleGAN is trained to transform contrast CT images into non-contrast images. This study can reduce manual segmentation effort and cost in CT imaging.

In the natural scene, using GANs to perform data augmentation also shows good results. In [6], the semantic segmentation problem is carried out in the natural scene, where data augmentation approach is used to balance the label distribution in order to improve segmentation performance. A Pix2pix HD [30] model is applied to generate realistic images on the condition of the specific semantic label map. The generated images not only improve segmentation performance of those classes with low accuracy, but also obtain a 1.3% to 2.1% increase in average segmentation accuracy.

Although GANs are powerful, it is very difficult to get high-resolution images and they require a huge amount of images to train. Thus, using GANs to solve small dataset problems still is a challenge [31].

4) THE APPLICATIONS AND HIGH-PERFORMANCE IMPLEMENTATIONS OF DEEP LEARNING IN BIOMEDICAL AND BIOLOGICAL INFORMATICS

At present, deep learning is widely used in biomedical and biological information domains. For example, [32], [33] use a genetic hierarchical network and SVM to predict credit score. References [34], [35] use a deep neural network to evaluate and diagnose the electrocardiogram signal. Reference [36] uses ResNet networks to recognize sensor signals. All these studies obtain promising results and show the advantages and high performance of deep learning. References [37], [38] present a CUDA-based CT image reconstruction tool based on the algebraic reconstruction technique (ART) or cuART. They propose a symmetry-based CSR format (SCSR) to further compress the CSR data structure and optimize data access for both Sparse Matrix-Vector (SpMV) and SpMV-T via a column-indices permutation, and the experimental results are pretty good. Reference [39] develops a system, MemXCT, that uses an optimized SpMV implementation with two-level pseudo-Hilbert ordering and multi-stage input buffering, which is a novel memory-centric approach that avoids redundant computations at the expense of additional memory complexity. MemXCT can reconstruct a large $(11K \times 11K)$ mouse brain tomogram in 10 seconds using 4096 KNL nodes (256K cores), which is a good result.

C. EM IMAGE CLASSIFICATION

For EM image classification, there are two basic categories of feature extraction techniques: "hand-crafted features" and "feature learning" [40]. The basic idea of "hand-crafted features" is as follows. First, we extract the features of EM images. Second, we classify these EM images according to the features. The hand-crafted features include global shape, local shape (including SIFT), texture, color, etc. However, hand-crafted features are insufficient for representing diverse appearances of EMs because hand-crafted features are manually designed based on prior knowledge and investigation. Compared to this, feature learning is a better technique for EM image classification, including Bag of Visual Words (BoVW), Sparse Coding (SC), deep learning, etc.

Different from extracting specific features, the technique of feature learning is to extract features that do not have meaning in practice but can represent diverse appearances of EMs in the hidden world. For example, in [40], conditional random fields and deep convolutional neural networks are used to extract features and classify EM images. Their dataset contains 20 classes of EMs. Each EM class is represented by 20 microscopic images. The mean average precision of the experimental results is 91.40%, and it shows that the classification accuracy is very high and the method is effective. In [41], SC is used to classify EM images. Their dataset

contains 15 classes of EMs. Each EM class is represented by 20 microscopic images. To overcome small dataset problem and effectively represent scarce training images, they use SC which extracts scient local features from an image and reconstructs it by a sparse linear combination of bases. They also use weakly supervised learning to jointly perform the localization and classification of EMs by examining the local information in training images. The mean average precision of the experimental results is about 55% and shows the effectiveness of the method.

As in the previous studies, EM image classification has always suffered from the problem of small datasets. To improve the classification accuracy, these studies all design better classification methods, rather than augment EM image datasets. So, EM image augmentation is a perspective of improving classification accuracy.

III. ENHANCED FRAMEWORK OF GENERATIVE ADVERSARIAL NETWORKS
A. GENERATIVE ADVERSARIAL NETWORKS
GAN is first introduced in [42], based on a game-theory that consists of a generator (G) network and a discriminator (D) network. The purpose of G is to generate fake images to deceive D, and the purpose of D is to discriminate real images and fake images generated by G. Through training, G and D are in a Nash equilibrium. To be specific, D cannot discriminate between real images and fake images generated by G. The framework of GAN is shown in Fig. 2.
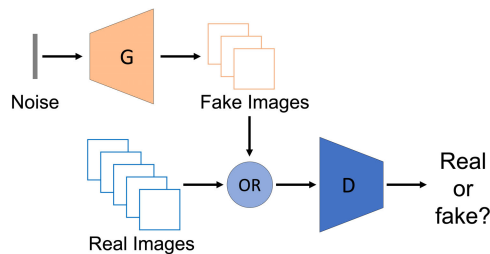

FIGURE 2. The framework of GAN.

In recent years, there are improved GANs because of good performance in generating images. For example, Deep Convolutional Generative Adversarial Networks (DCGAN) [43], Wasserstein Generative Adversarial Networks (WGAN) [44], and Improved Wasserstein Generative Adversarial Networks (WGAN-GP) [45] are highly praised GAN models.

The discriminator and generator of DCGAN use the convolutional neural network (CNN) to replace the multi-layer perceptron in GAN. Meanwhile, to make the whole network differentiable, the pooling layer in CNN is removed, and the global pooling layer is used to replace the full connection layer to reduce the computation. DCGAN is an improvement on GAN, and its modified version of the original GAN is mainly in network structure. Up to now, DCGAN has greatly improved the stability of GAN training and the resolution of the generated images.

Different from DCGAN, WGAN mainly improves GAN from the loss function. In theory, WGAN gives the reason

for the instability of GAN training, that is, the cross-entropy (JS divergence) [46] is not suitable for measuring the distance between the generated data distribution and the real data distribution. Instead, Wasserstein distance [44] is used to measure the distance between the generated data distribution and the real data distribution, which theoretically solves the problem of unstable training. However, the use of the Wasserstein distance requires Lipschitz continuity [47]. To satisfy this condition, the authors impose the Lipschitz continuity by restricting the weights to a range, but this also causes the problem of gradient disappearance or gradient explosion.

To solve the problem of gradient disappearance or gradient explosion, and to find an appropriate way to meet the continuity condition of Lipschitz, in WGAN-GP, the author proposed the method of gradient penalty to meet the continuity condition. WGAN-GP has a faster convergence rate than WGAN and can produce higher resolution images.

B. DETAILS OF THE EF-GANs
As shown in Fig. 3, the structure of the EF-GANs is as follows: In step-1, we rotate original images in the same direction and perform color space transformation. In (I), let the pixel value of the GT image at the point (x, y) be $I_{GT}(x, y)$. In (II), the GT image is rotated every $3°$ counterclockwise from $0°$ to $180°$. When the area of the bounding rectangle is the smallest, the bounding rectangle is called the minimum bounding rectangle. When the radian of rotation angle is θ, the centroid of GT image is $C = (x_c, y_c)$, where

$$\begin{cases} x_c = \dfrac{\sum\limits_{x=1}^{n}\sum\limits_{y=1}^{m} x \cdot I(x, y)}{\sum\limits_{x=1}^{n}\sum\limits_{y=1}^{m} I(x, y)} \\ y_c = \dfrac{\sum\limits_{x=1}^{n}\sum\limits_{y=1}^{m} y \cdot I(x, y)}{\sum\limits_{x=1}^{n}\sum\limits_{y=1}^{m} I(x, y)}. \end{cases} \quad (1)$$

The EF-GANs combine geometric transformation methods and GANs for EM image augmentation. It is not a new GAN but an enhanced framework of GANs. Various GANs can be inserted into the EF-GAN framework. As shown in Fig. 4, the radian of rotation angle is θ. The long axis is the line when the midpoint of the two short sides of the bounding rectangle is connected. The short axis is the line that the midpoint of the two long sides of the bounding rectangle is connected. They are two blue dashed lines that are perpendicular to each other, and the orange cross is the centroid.

In (III), we regard the MBR and centroid of the GT image as an original image because of their corresponding relationship. In (IV), we rotate the original image to make its long axis horizontal and centroid on the same side. Specifically, if the centroid is on the left of the short axis, we rotate the original image by θ, shown in Fig. 5(a). If the centroid is on the right of
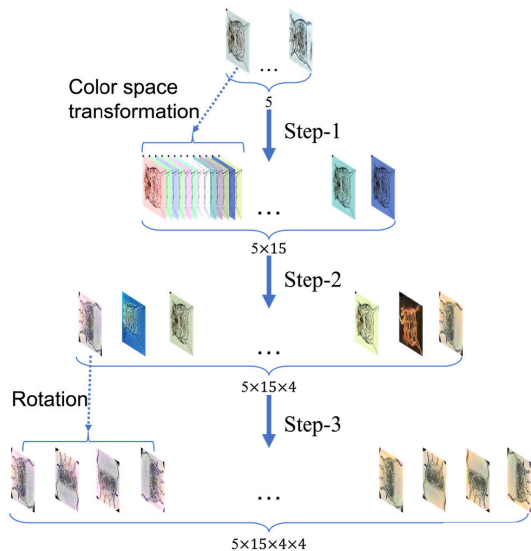
H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

IEEE_Access_

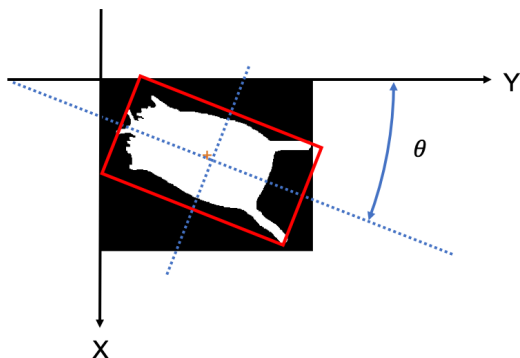**FIGURE 3.** The structure of the EF-GANs.

FIGURE 4. An example of MBR for an EM. The MBR is marked with a red box.
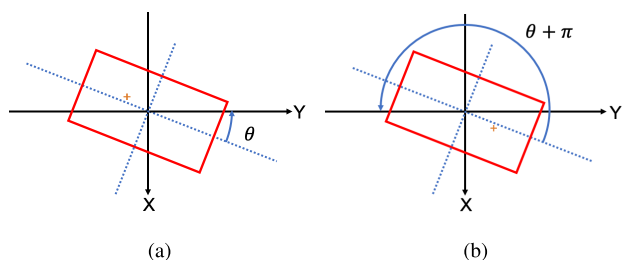

FIGURE 5. Two different rotation angles. 5(a) is the image that centroid is on the left. 5(b) is the image that centroid is on the right.

the short axis, we rotate the original image by $\theta + \pi$, which is shown in Fig. 5(b). In (V), we use the image to perform color space transformation. Specifically, we transform the image into fifteen color spaces by multiplying a random number between 0 and 1. After step-1, we augment the original image to fifteen times.

In step-2, the augmented images from step-1 are input into GANs to generate more images. Specifically, we choose various GANs and insert them into the EF-GANs, such as DCGAN, WGAN, and WGAN-GP.

In step-3, because the angle of an EM in an image does not affect its class label, we rotate images generated by the EF-GANs in different directions to increase image diversity.

As shown in Fig. 3, in step-1, we augment images fifteen times. This can solve the problem that GANs need plenty of images to train nets so that they can generate high resolution images. In step-2, we use GANs to augment images further and increase image diversity. In step-3, we enhance the image direction feature through rotation. After step-1, step-2, and step-3, we augment images 240 times and this method does not affect image quality. Because we rotate EM images to make their long axis horizontal and centroid on the same side, we help the GAN reduce the distance between the real images and the fake images, making it easier to generate the images. Therefore, GANs can generate effective images. Although some background information of EM images would be lost after step-1, the background information contains less information about GANs image generation and EM classification. So it is worth it.

IV. EXPERIMENTS
A. EXPERIMENTAL SETTING
1) IMAGE DATASETS
In our work, we use the Environmental Microorganism Data Set 5th Version (EMDS-5) [9], containing 21 classes of EMs $\{\omega_1, \ldots, \omega_{21}\}$ as shown in Fig. 6. Each EM class is represented by 20 microscopic images and their corresponding ground truth images, thus EMDS-5 includes altogether 420 scenes.

2) DATASETS SETTING
We randomly divide each EM class into three parts with a ratio of 1:1:2, corresponding to the training, validation, and test sets. Therefore, for each EM class, we have five, five, and ten images for training, validation, and test. To avoid interference, we only use the training dataset for image augmentation, and the validation and test sets are used for the evaluation. In step-1, five training images are expanded to 75 images in each EM class. Specifically, we multiply the EM images by a random number between 0 and 1, and map the EM images to 15 different color spaces. In step-2, we use these 75 images to train EF-GANs to get 300 generated images. In the experiment, we observe that after generating 240 to 250 images, the images become more and more similar, so we heuristically cut off the generated images at 300 to keep away from redundant information. In step-3, we augment these 300 images with rotation by 0°, 90°, 180°, and 270°, which results in 1200 images for training.

3) EXPERIMENTAL ENVIRONMENT
The experiment is conducted by Python 3. The models we used in our experiment are implemented using Keras [48] framework with Tensorflow [49] as backend.
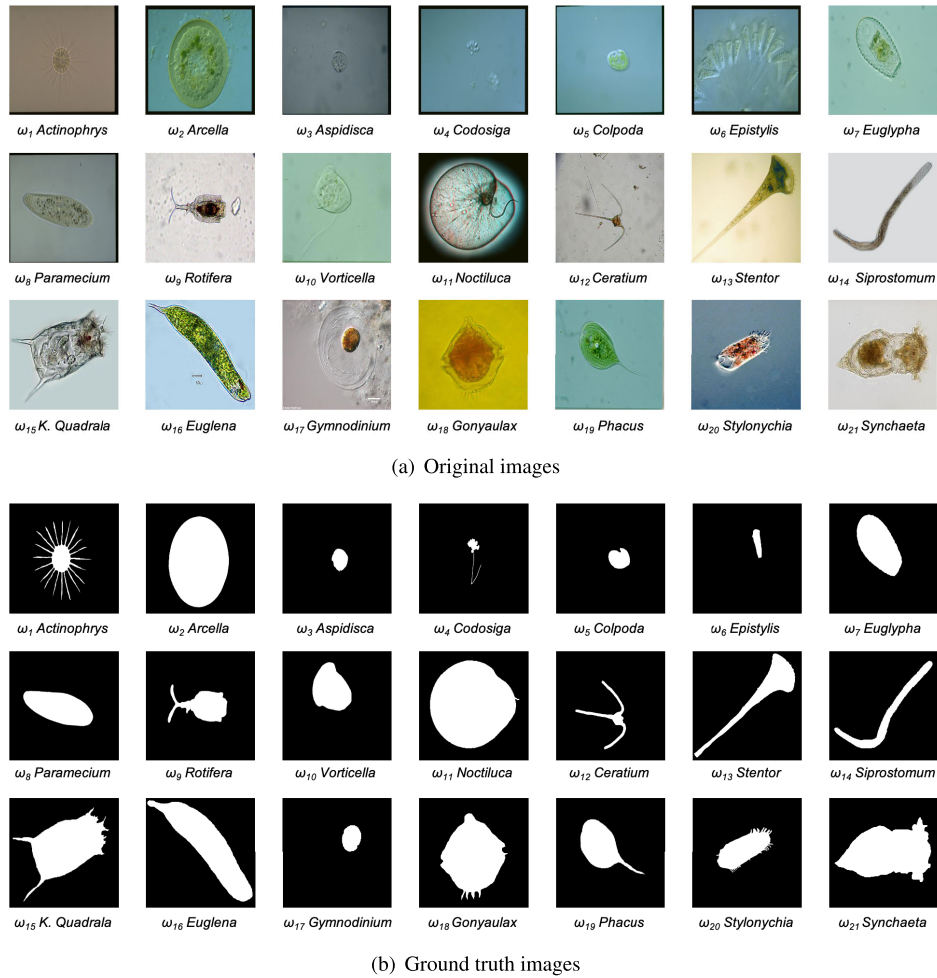
IEEE *Access*

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

(a) Original images

(b) Ground truth images

FIGURE 6. Examples of original and ground truth images in EMDS5.
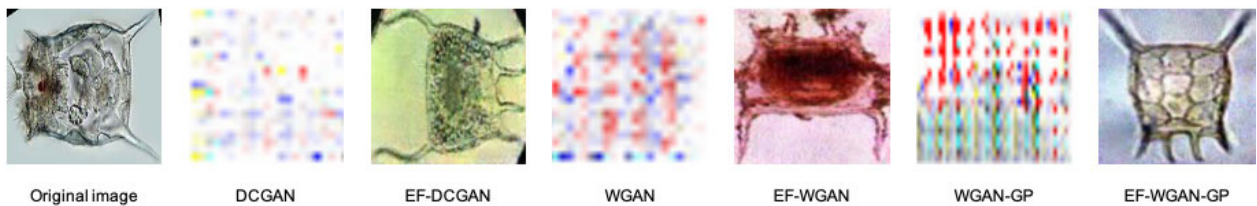

FIGURE 7. A comparison between images generated by original GANs and their corresponding EF-GANs.

In our experiment, we use a workstation with Windows 10 operating system with Intel(R) Core(TM) i7-8700K CPU with 3.70GHz, 16GB RAM, and NVIDIA GEFORCE RTX 1080 8GB.

B. EVALUATION OF IMAGE GENERATION

As shown in Fig.7, there is a comparison between the images generated by original GANs (i.e., DCGAN, WGAN, WGAN-GP) and EF-GANs (i.e., EF-DCGAN, EF-WGAN, EF-WGAN-GP). We can see that the EM images generated by EF-GANs have higher visual quality than that generated by the original GANs. To be specific, the images

generated by original GANs are full of noise, so it is impossible to distinguish what EMs they are. In contrast, EF-GANs cannot only generate distinguishable EM images but also have accepted image resolution. The reason why EF-GANs perform better than the original GANs in generating EM images is that the EF-GANs can generate rotation-invariant images, but the original GANs cannot. For example, in [43], face images are augmented using DCGAN, while face images are directional (eyes are up, nose is in the middle, and mouth is down.). In [44], indoor images are generated by WGAN, while indoor images are directional (furniture is on the floor). In [45], WGAN-GP generates bedroom images,
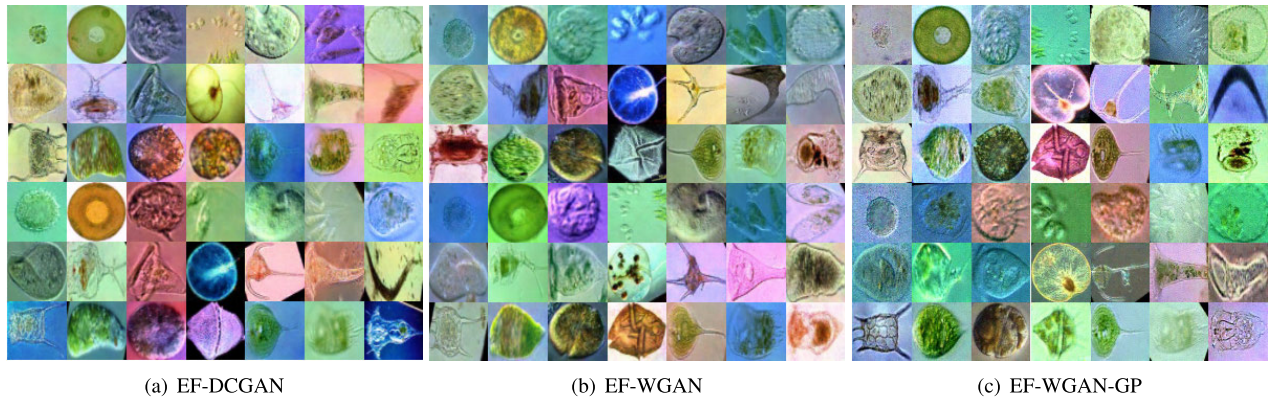
H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

IEEE *Access*

(a) EF-DCGAN   (b) EF-WGAN   (c) EF-WGAN-GP

FIGURE 8. An example of augmented images genetated by EF-GANs.

while bedroom images are directional (beds and chairs are on the floor). EF-GAN unifies EM direction by using classical geometric methods, which makes the generation of EM images take no account of orientation, so that high-quality EM images can be generated. There are differences in EM image generation between different EF-GANs. To be specific, the EM images generated by EF-DCGAN are artificially distinguished and have pure backgrounds. However, some images are fuzzy. The resolution of EM images generated by EF-DCGAN are not high and their outlines are not very clear. The EM images generated by EF-WGAN have clearer outlines, purer backgrounds, and higher resolution. The EM images generated by EF-WGAN-GP have clearer outlines, higher sharpness, and higher resolution and less noise. An example of the augmented images generated by EF-DCGAN, EF-WGAN, and EF-WGAN-GP is shown in Fig. 8.

C. EVALUATION OF DATA AUGMENTATION FOR EM IMAGE CLASSIFICATION

1) AVERAGE PRECISION (AP) AND MEAN AVERAGE Precision (mAP)

We set the class of samples as K and the total number of each class of samples as N, including M positive samples. For each class, we use the trained model to obtain the test samples' confidence scores and sort them in descending order. According to the confidence scores and labels of all test samples, we get the classification results are True Positive (TP), True Negative (TN), False Positive (FP), or False Negative (FN).

According to Eq. 2 3, the precision and recall from top-1 to top-N are calculated. Then there can be M recall values from top-1 to top-N, which are $1/M, 2/M, \ldots, M/M$, respectively. For each recall value, we can calculate a maximum precision. We average M maximum precision values to get AP in Eq. 4.

There are K classes of EM images in all. According to Eq.5, we average all classes of APs to get mAP.

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

$$\text{AP} = \frac{1}{M} \sum_{i=1}^{M} \text{Precision}_{max}(i) \tag{4}$$

$$\text{mAP} = \frac{1}{K} \sum_{j=1}^{K} \text{AP}(j) \tag{5}$$

2) EVALUATION OF DATA AUGMENTATION FOR EM IMAGE CLASSIFICATION

In this section, we demonstrate that image augmentation improves the EM image classification performance. We choose the VGG16 networks to test the classifier's improvement by adding different numbers of augmented images to the training set. Especially, in this multi-class classification task, we use a one-vs-rest strategy to design our experiment. Firstly, for 21 EM image categories $\{\omega_1, \ldots, \omega_{21}\}$, we divide them into positive and negative classes. ω_i, for $i \in 1, \ldots, 21$ is the positive class, and $\bigcup_{j=1:21; j \neq i} \omega_j$, for $i \in 1, \ldots, 21$ is the negative class. Secondly, for each one-vs-rest classification task, we add different numbers of images generated by different EF-GANs to the training set. We use the training and validation sets to train the VGG16 networks and calculate the Average Precision (AP). Thirdly, we calculate the mean Average Precision (mAP) of each one-vs-rest classification task. Finally, we calculate the mAP of each task. The result is shown in Table. 1.

For each one-vs-rest classification task, we add different numbers of images generated by different EF-GANs to the training set. For example, ''EF-DCGAN,15'' means adding 15 images generated by EF-DCGAN to the training set. ''Original'' means only using the original training set.

Because of the extremely unbalance of our datasets, only using accuracy to measure the effect of classification is not enough. The ratio of positive classes to negative classes in our validation sets is 1:20. If the classifier always predicts a negative example, the accuracy is 95.24%, which is very high. The TN is 100% but the FN is 0%. So this result

IEEE *Access*

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

TABLE 1. A comparison of the EM image classification results using the VGG16 networks with original and augmented training sets in percentage (%). The first row shows different classification tasks. The second to the second bottom rows show the classification APs. The first column shows the EF-GANs and the numbers of added images. The bold values for each column are more significant than "Original" or the same as "Original," and the values with stars are the maximum for each column.

	ω_1 vs rest	ω_2 vs rest	ω_3 vs rest	ω_4 vs rest	ω_5 vs rest	ω_6 vs rest	ω_7 vs rest	ω_8 vs rest	ω_9 vs rest	ω_{10} vs rest	ω_{11} vs rest
Original	69.56	85.56	74.54	43.17	28.09	*87.62	67.22	6.17	4.74	84.55	50.12
EF-DCGAN,15	76.43	96.67	85.56	22.34	31.51	54.31	63.67	70.76	48.74	63.62	84.17
EF-DCGAN,35	70.00	88.50	79.33	49.27	23.69	47.92	64.27	57.45	45.04	66.77	58.83
EF-DCGAN,55	63.62	*100.00	45.97	41.31	35.08	50.79	63.43	62.22	51.33	54.51	64.82
EF-DCGAN,75	71.67	*100.00	71.67	43.17	23.80	51.67	65.33	62.26	45.14	60.56	57.94
EF-DCGAN,95	69.56	92.50	42.76	41.15	24.87	50.32	52.83	64.17	30.52	54.67	63.12
EF-WGAN,15	*86.00	*100.00	60.00	*66.33	32.82	67.00	*68.94	58.10	56.72	55.48	74.25
EF-WGAN,35	63.42	*100.00	75.00	38.29	33.09	50.57	55.67	68.98	63.36	49.33	71.98
EF-WGAN,55	82.11	87.67	70.14	28.28	16.57	55.33	67.04	78.93	68.50	54.67	71.43
EF-WGAN,75	60.00	*100.00	69.89	48.10	35.00	53.01	58.85	74.54	*72.00	58.00	81.56
EF-WGAN,95	72.48	*100.00	60.89	35.79	22.94	47.09	68.29	79.33	65.44	56.60	79.12
EF-WGAN-GP,15	57.45	92.67	83.14	41.01	*37.87	55.31	48.73	*81.56	51.67	83.13	77.69
EF-WGAN-GP,35	55.40	*100.00	*94.29	37.82	31.70	47.09	52.14	69.78	51.23	*85.00	81.00
EF-WGAN-GP,55	63.94	90.00	91.11	53.81	32.83	57.26	54.14	52.45	42.82	73.89	86.25
EF-WGAN-GP,75	50.99	92.67	68.81	29.80	24.61	49.72	55.01	56.53	43.78	56.85	*96.67
EF-WGAN-GP,95	46.61	87.62	84.33	46.44	29.28	48.49	54.68	78.14	68.89	67.56	79.62

|	ω_{12} vs rest	ω_{13} vs rest	ω_{14} vs rest	ω_{15} vs rest	ω_{16} vs rest	ω_{17} vs rest	ω_{18} vs rest	ω_{19} vs rest	ω_{20} vs rest	ω_{21} vs rest	mAP
Original	4.52	27.48	4.34	45.24	4.23	82.56	37.00	4.81	56.99	3.23	41.51
EF-DCGAN,15	35.35	43.06	49.94	77.22	36.61	70.87	68.68	45.28	62.99	27.99	57.89
EF-DCGAN,35	50.11	36.50	46.35	76.33	38.96	75.93	78.60	42.31	74.33	42.13	57.74
EF-DCGAN,55	16.83	49.22	51.32	*90.29	42.93	60.52	78.60	46.72	72.78	39.55	56.28
EF-DCGAN,75	28.89	47.21	65.40	67.40	53.56	83.93	79.76	35.06	68.33	37.26	58.10
EF-DCGAN,95	33.82	*74.33	42.05	64.35	34.21	81.00	*92.67	40.48	68.33	28.63	54.59
EF-WGAN,15	*57.26	64.17	47.60	68.33	*88.33	83.93	83.33	42.86	63.43	20.08	*64.05
EF-WGAN,35	47.62	46.57	63.36	73.33	37.78	71.12	55.33	34.59	52.74	20.99	55.86
EF-WGAN,55	46.76	64.67	36.26	87.69	37.12	*87.62	70.33	42.18	59.03	*54.99	60.35
EF-WGAN,75	46.21	49.13	*68.56	63.55	53.94	78.93	59.63	32.03	53.43	23.41	59.04
EF-WGAN,95	39.61	47.24	31.76	58.16	58.25	85.29	71.96	35.01	59.50	27.49	57.25
EF-WGAN-GP,15	43.67	72.03	55.04	77.42	60.79	81.00	40.60	45.93	67.67	26.88	61.01
EF-WGAN-GP,35	39.84	59.78	37.60	75.14	63.10	76.88	51.86	*51.11	66.36	29.03	59.82
EF-WGAN-GP,55	37.42	69.78	36.21	74.54	43.50	84.76	45.54	43.02	*81.43	24.09	58.99
EF-WGAN-GP,75	57.26	66.01	49.75	76.11	53.66	61.90	58.59	39.47	57.12	24.26	55.69
EF-WGAN-GP,95	47.89	48.33	57.98	72.31	60.15	83.50	70.17	25.77	73.93	17.89	59.50
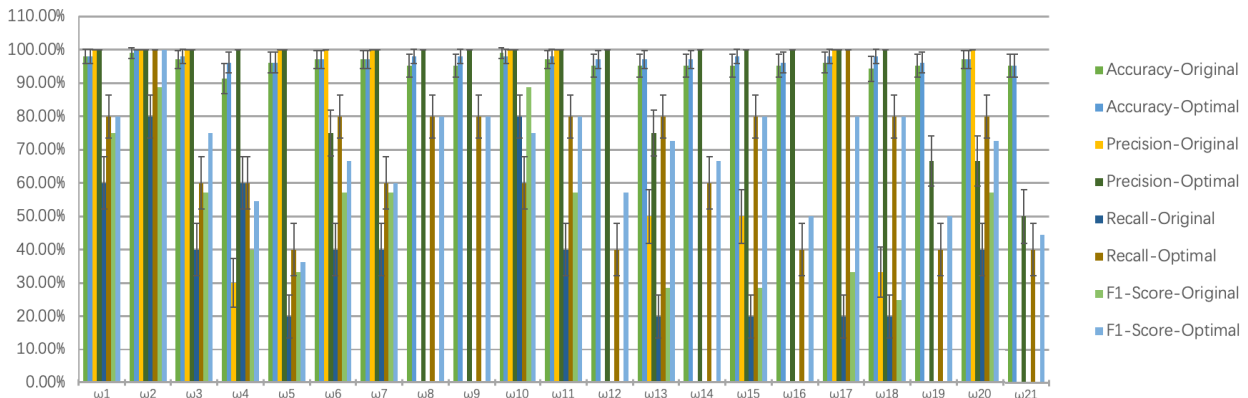
FIGURE 9. The Accuracy, Precision, and Recall of the EM classification results with 90% confidence interval, and F1 score of the EM classification results. Comparison between the original training sets and the augmented training sets of the optimal of EF-GANs. The horizontal axis represents various EM classes. The vertical axis represents the value of Accuracy, Precision, Recall and F1 score.

is meaningless. Therefore, we use precision and recall with confidence intervals, as well F1 score to measure the effect of our networks. F1 score is the harmonic mean of precision and recall. As shown in the Fig. 9, we calculate the accuracy, precision, recall, and F1 score. There are 20 of 21 tasks that the accuracy of "Optimal" is greater than or equal to the accuracy of "original," 19 of 21 tasks that the precision of "Optimal" is greater than or equal to the precision of "original," 20 of 21 tasks that the recall of "Optimal" is greater than or equal to the recall of "original," and 20 of 21 tasks that the F1 score of "Optimal" is greater than or

equal to the F1 score of "original." The 90% confidence interval for all these results are between 0% and 8%. The results show that the EM images augmented by EF-GANs can improve the performance of classification.

3) DISCUSSION OF THE VGG16 CLASSIFICATION RESULTS
We evaluate the ability of EM image generation of EF-GANs by the improvement of VGG16 classification results. Table. 1 shows the classification performance of VGG16 networks. There are 21 one-vs-rest EM image classification tasks. After data augmentation, the APs with VGG16 networks increase
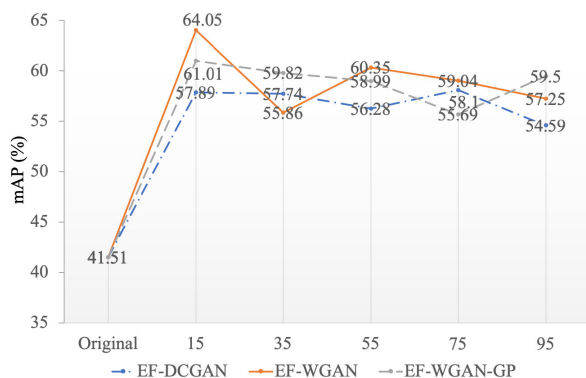
H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

IEEE *Access*

FIGURE 10. A comparison of mAPs of adding different numbers of EM images augmented by EF-DCGAN, EF-WGAN, and EF-WGAN-GP into training sets (%). The horizontal axis represents the number of augmented EM images and the vertical axis represents the mAP of classification results of VGG16 networks.

between 4.5% and 84.1% on 20 tasks. "EF-WGAN,15" has the highest AP with VGG16 networks in six tasks and its mAP is also the highest, showing that VGG16 should increase the classification performance with a small number of augmented images. That is because when a large number of images are added, the newly augmented images will have background information on other classes of images. The mAP results in Table. 1 are represented by a line graph in Fig. 10. We can see that the mAP is the highest when we add 15 images augmented by EF-WGAN and EF-WGAN-GP, and the mAP is the second highest when we add 15 images augmented by DCGAN. It shows that adding a small number of EM images into training sets have better results. In our model, we add different numbers of fake EM images generated by EF-GANs into training sets. When we add a small number of EM images into training sets, the mAP increase. However, when we add more fake EM images generated by EF-GANs into training sets, it brings much noise as well, as a result of the decrease of mAP. It shows that adding more EM images into training sets is not always helpful because the generated EM images are fake. In conclusion, if there are many fake EM images in the training sets, it will bring more noise, and when the noise accumulates to a certain extent, the training effect will decline.

In the VGG16 classification tasks of "ω_5 vs rest," "ω_7 vs rest," "ω_{10} vs rest," and "ω_{17} vs rest," the increase of results is less than 10%. By observing the information of these EMs, we find that the original results are high over 67% in the task of "ω_7 vs rest," "ω_{10} vs rest" and "ω_{17} vs rest," so that the increase of results is less than 10%. The boundary of "ω_5" augmented images is not clear, so that the increase of its result is less than 10%.

In the VGG16 classification tasks of "ω_8 vs rest," "ω_9 vs rest," "ω_{11} vs rest," "ω_{12} vs rest," "ω_{13} vs rest," "ω_{14} vs rest," "ω_{15} vs rest," "ω_{16} vs rest," "ω_{18} vs rest," "ω_{19} vs rest," and "ω_{21} vs rest," the increase of results is more than 40%. The shapes of these EMs are like a circle and some

have a few elongated body parts so that the increase of results is more than 40%. In the task of "ω_6 vs rest," it has a better result that does not add augmented images to the training set. By observing the information of EM "ω_6," we found that it is difficult to classify because of its characteristic that some of them are clustered together and some are isolated. In conclusion, EF-GANs have a strong ability to generate independent EMs, but a weak ability to generate a cluster of EMs. Adding a small number of generated images have a better effect than adding a large number of generated images.

D. ADDITIONAL EXPERIMENT A: TEST WITH ResNet50

To improve the robustness of the model, we have implemented the ResNet50 networks to complete 21 one-vs-rest EM image classification tasks and the result is shown in Table. 2. The APs value with ResNet50 networks increase between 8.7% and 38.7% in 12 tasks, decline slightly in 7 tasks, and stay the same and equal 100% in 2 tasks. This proves the effectiveness of our method. However, there are seven classes of EM images where the APs values are decreasing. By observing these images, we found that the boundaries of these EM images are not clear and almost mixed with the background so that the increase of results is less than 10%. So, our method has a good effect on the generation of single EM with the clear edge and pure background, but a little weak for the generation of EM images with the fuzzy edge, impure background, or clusters.

E. ADDITIONAL EXPERIMENT B: TEST WITH OTHER DATASET

As the EMDS-5 test-set has only 10 images in each class of EM, in order to further prove our method, we collect another 630 EM images for testing. There are 21 classes of EM images, and each class of EM images has 30, a total of 630. According to the test results of EMDS-5 in Table 1, we choose three kinds of VGG16 networks to test each class of EM images. The first is the VGG16 networks without augmentation, "Original." The second is the VGG16 networks with the optimal augmented setting for all EM classes, "EF-WGAN, 15" from the previous experimental result in Table 1. The third is the VGG16 networks with the optimal augmented setting for single EM classes, which is 21 different optimal VGG16 networks for each single class of EM images from the previous experimental result in Table 1. Therefore, we can reflect the generalization performance of our method and the result is shown in Table 3.

From Table 3, we can find that in the 21 one-vs-rest EM image classification tasks, there are 20 tasks that the APs with "Optimal setting for all EM classes" and "Optimal setting for single EM classes" are 1.66% to 88.1% higher than "Original." There is only one task that the APs with "Original" is higher than "Optimal setting for all EM classes" and "Optimal setting for single EM classes." The mAP with "Optimal setting for all EM classes" and "Optimal setting for single EM classes" are 31.95% and 30.81% higher than "Original," respectively. It shows that our generated images

IEEE *Access*

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

TABLE 2. A comparison of the EM image classification results using ResNet50 networks with original and augmented training sets. (In [%].) The first row shows different classification tasks. The second to the bottom rows show the classification APs. The first column shows the EF-GANs and the numbers of added images. The values in bold for each column are more significant than "Original" or the same as "Original," and the values with stars are the maximum for each column.

Class	ω_1 vs rest	ω_2 vs rest	ω_3 vs rest	ω_4 vs rest	ω_5 vs rest	ω_6 vs rest	ω_7 vs rest	ω_8 vs rest	ω_9 vs rest	ω_{10} vs rest	ω_{11} vs rest
Original	89.09	*100.00	*94.29	41.84	*45.52	*100.00	32.40	*85.26	41.78	30.63	83.93
EF-DCGAN,15	*100.00	*100.00	63.93	39.63	16.07	92.67	17.32	65.94	58.88	19.37	68.69
EF-DCGAN,35	87.11	90.29	77.22	47.12	16.95	*100.00	38.62	75.61	46.57	27.39	*92.67
EF-DCGAN,55	*100.00	*100.00	92.67	34.43	7.75	78.62	21.97	47.64	41.67	32.97	81.00
EF-DCGAN,75	*100.00	*100.00	81.67	39.11	20.03	94.29	19.08	62.00	51.46	26.68	87.67
EF-DCGAN,95	*100.00	92.67	68.10	53.59	25.34	96.67	13.86	62.48	53.67	34.65	76.44
EF-WGAN,15	*100.00	96.67	78.93	36.22	27.60	79.62	*41.37	70.60	*61.48	27.14	68.57
EF-WGAN,35	88.50	81.00	81.43	56.25	20.54	96.67	36.02	72.48	21.80	37.06	82.54
EF-WGAN,55	81.88	92.67	65.33	44.55	28.89	96.67	38.44	67.83	46.93	37.10	80.09
EF-WGAN,75	79.12	92.67	79.09	41.17	28.81	87.67	17.19	71.03	35.38	22.44	87.67
EF-WGAN,95	81.03	92.67	60.54	26.74	20.91	71.00	21.75	75.00	29.45	18.64	80.09
EF-WGAN-GP,15	*100.00	*100.00	69.09	42.51	26.17	68.62	26.76	63.90	52.67	*39.33	82.67
EF-WGAN-GP,35	94.29	92.67	54.22	44.26	19.96	82.62	31.62	83.69	53.57	16.55	87.11
EF-WGAN-GP,55	88.50	96.67	72.54	*62.71	20.14	78.62	21.99	59.67	26.33	37.86	75.00
EF-WGAN-GP,75	81.03	91.11	50.89	56.06	21.71	78.62	18.00	65.81	30.38	30.61	*92.67
EF-WGAN-GP,95	71.03	*100.00	57.22	35.92	23.90	96.67	10.11	61.83	28.06	22.29	87.67

Class	ω_{12} vs rest	ω_{13} vs rest	ω_{14} vs rest	ω_{15} vs rest	ω_{16} vs rest	ω_{17} vs rest	ω_{18} vs rest	ω_{19} vs rest	ω_{20} vs rest	ω_{21} vs rest	mAP
Original	72.50	90.00	63.03	*85.29	*79.33	27.36	22.05	*67.79	41.35	*76.44	*65.23
EF-DCGAN,15	60.57	87.69	76.26	57.56	50.88	28.69	*60.75	13.28	27.02	49.67	54.99
EF-DCGAN,35	49.55	94.29	83.50	64.74	37.56	37.10	40.84	29.31	29.35	35.46	57.20
EF-DCGAN,55	59.90	96.67	83.93	58.06	37.46	55.17	27.42	24.85	39.51	52.15	55.90
EF-DCGAN,75	57.82	62.52	85.83	59.56	37.99	29.92	41.64	17.36	30.03	34.25	54.23
EF-DCGAN,95	49.70	87.11	87.11	57.17	41.06	*58.04	35.58	14.94	38.86	25.50	55.84
EF-WGAN,15	66.92	90.29	73.43	67.22	66.83	28.55	42.18	11.13	31.70	38.94	57.40
EF-WGAN,35	*79.85	*100.00	67.46	69.78	67.00	41.22	33.05	6.36	40.27	36.24	57.88
EF-WGAN,55	67.77	90.29	81.43	62.99	58.67	45.78	41.46	13.84	*55.38	40.22	58.96
EF-WGAN,75	54.38	90.29	71.43	72.33	51.12	55.66	42.67	14.58	54.22	40.33	56.63
EF-WGAN,95	61.16	80.83	85.09	60.32	47.56	47.00	26.92	35.25	52.58	50.36	53.57
EF-WGAN-GP,15	71.33	92.67	86.00	68.19	62.95	46.03	46.33	9.61	28.06	36.25	58.05
EF-WGAN-GP,35	73.75	83.69	63.94	67.26	52.22	46.89	40.10	24.79	50.47	39.13	57.28
EF-WGAN-GP,55	48.24	73.42	87.11	56.08	44.03	49.58	47.27	4.42	30.95	47.21	53.73
EF-WGAN-GP,75	36.61	85.29	*91.11	76.27	23.36	33.04	27.97	10.66	28.95	33.83	50.67
EF-WGAN-GP,95	34.48	96.67	76.67	69.04	43.81	21.62	45.48	18.28	26.76	27.30	50.23

TABLE 3. A comparison of the EM image classification results using VGG16 networks with original and best augmented settings. (In [%].) The first row shows different classification tasks. The second row shows the classification APs without augmentation. The third row shows the classification APs with the optimal augmented setting for all EM classes (EF-WGAN,15). The fourth row shows the classification APs with the optimal augmented setting for single EM classes. The values in bold are the maximum for each column.

	ω_1 vs rest	ω_2 vs rest	ω_3 vs rest	ω_4 vs rest	ω_5 vs rest	ω_6 vs rest	ω_7 vs rest	ω_8 vs rest	ω_9 vs rest	ω_{10} vs rest	ω_{11} vs rest
Original	93.33	59.25	31.75	40.56	15.84	77.57	58.57	8.80	9.71	65.54	80.00
Optimal setting for all EM classes (EF-WGAN,15)	94.99	74.61	48.97	59.33	30.29	75.03	66.59	83.52	51.52	73.45	95.94
Optimal setting for single EM classes	94.99	74.61	44.56	59.33	31.1	75.03	66.59	88.45	46.16	59.89	94.14

	ω_{12} vs rest	ω_{13} vs rest	ω_{14} vs rest	ω_{15} vs rest	ω_{16} vs rest	ω_{17} vs rest	ω_{18} vs rest	ω_{19} vs rest	ω_{20} vs rest	ω_{21} vs rest	mAP
Original	10.73	36.84	9.07	64.57	7.71	50.43	45.35	10.43	43.6	9.07	39.46
Optimal setting for all EM classes (EF-WGAN,15)	75.1	82.75	90.77	73.12	90.68	54.03	70.87	85.41	68.64	54.13	71.42
Optimal setting for single EM classes	75.1	75.43	97.17	74.13	90.68	58.59	68.11	87.18	62.2	52.22	70.27

are useful for EM image classification. For the only one task with reduced AP value, the AP decrease by less than 10% compared with without augmentation. By observing that class of EM images, we find that they are all clustered together so that the VGG16 networks cannot distinguish. It shows that our method has weakly ability for generating clustered EMs to increase the classification APs. In conclusion, we use a completely different EM test-sets to test the VGG16 networks with promising results. It shows that our augmented EM images are not only useful for EM image classification but also prove the robustness of our method.

F. COMPUTATION TIME

In terms of computation time, it takes about 20-30 minutes to train DCGAN, WGAN, and WGN-GP with batchsize 32, stable learning rate 0.0002, and 800 training epochs for generating a single class of EM images. The selected parameters are the same as that in [44] to train the EF-GANs better. For classifying the single class of EM images, we take about 65 seconds to train VGG16 and ResNet50 with batchsize 5, stable learning rate 0.001, and 50 training epochs, and 13 seconds to test. Considering that the performance of our GPU, NVIDIA GEFORCE RTX 1080 8GB, is not superior enough, if we switch to NVIDIA GEFORCE RTX 2080Ti 11GB [50], the training time can be reduced by 40%, theoretically.

V. CONCLUSION AND FUTURE WORK

In this paper, we propose an EF-GANs network for the EM image augmentation task. In step-1, we rotate original images to the main axis direction and perform color space transformation. In step-2, we add the augmented images from step-1 to EF-GANs to generate images. In step-3, we rotate the generated images from step-2 to different directions. Finally, we augment images 300 times. We solve the problem that a small dataset cannot use GANs to perform image

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

IEEE *Access*

augmentation. After that, we add the augmented images to the training set of VGG16 networks to test their validity, which in the 21 one-vs-rest EM image classification tasks, the APs of 20 EM categories are improved and among them 11 categories are over 40%. It shows that our method can augment high quality EM images and can improve the results of EM image classification. There is only one class of EM images that the APs with VGG16 decrease. By observing this class of EM images, we find that some of EMs are clustered together, so they are hard to be generated and the APs with VGG16 are declining. It shows that EF-GANs have a weak ability to generate EM images in which multi EMs are clustered together.

To test the quality of the generated EM images, we also use the ResNet50 networks to complete 21 one-vs-rest EM image classification tasks under the same experimental setting as VGG16 networks. The result shows that in the 21 one-vs-rest EM image classification tasks, the APs with ResNet50 networks increase between 8.7% and 38.7% in 12 tasks, decline slightly in 7 tasks, and stay the same and equal 100% in 2 tasks. The result shows that we can augment high quality EM images and increase the result of classification. By observing these seven classes of EM images where the results decrease, we find that the boundaries of these EM images are not clear and almost mixed with the background so that we cannot augment EM images with high resolution.

To further prove our method, we use another totally different EM image set for testing. We use three previous VGG16 networks that training by EMDS-5, which are "Original," "EFWGAN,15" and the optimal VGG16 networks for each single class of EM images. The result shows that in the 21 one-vs-rest EM image classification tasks, there are 20 tasks that the APs with "Optimal setting for all EM classes" and "Optimal setting for single EM classes" are higher than "Original." It shows that our augmented images have high quality and high resolution, and they can improve the APs of EM image classification. Considering that we use different EM images to test, it also shows that EF-GANs have good robustness for other EM image datasets.

In conclusion, EF-GANs have a strong ability to augment EM images which are single EMs and have clear boundaries, but a weak ability to augment EM images that EMs are clustered together and have fuzzy boundaries. And EF-GANs have strong generalization performance for other EM image datasets. In the future, we plan to test our EF-GANs on more GANs to find a more effective solution to augment the limited rotation-invariant EM images. We also plan to combine other geometric transformation techniques with GANs to augment EM images. For example, in [51], radial transformation is used in medical image augmentation for training deep neural networks. We will combine radial transformation with GANs to augment EM images. We will use the technique of GPU parallel acceleration [52] to further improve the training time of EF-GANs. And we will use our EF-GANs on other medical images such as cervical histopathology images, CT images of lung cancer, and CT-MR brain images.

In addition, in [53], a new GAN called GauGAN is proposed for image-to-image translation. It can convert a semantic segmentation mask to a photorealistic image and has spatially-adaptive normalization, a simple but effective layer for synthesizing photorealistic images given an input semantic layout. The GauGAN can produce photorealistic outputs for diverse scenes including indoor, outdoor, landscape, and street scenes. Hence, we plan to use this model for EM image augmentation, where the GT images of EM will be fed as the semantic layout to the GauGAN, and then generate the EM images. It is obvious that synthesising GT images of EM is easier than synthesizing EM images. So, we can synthesize a number of GT images of EMs, then use GauGAN to generate more EM images.

REFERENCES

[1] I. Pepper and C. Gerba, "Aeromicrobiology," in *Environmental Microbiology*. Amsterdam, The Netherlands: Elsevier, 2015, pp. 89–110.

[2] C. Li, K. Wang, and N. Xu, "A survey for the applications of content-based microscopic image analysis in microorganism classification domains," *Artif. Intell. Rev.*, vol. 51, no. 4, pp. 577–646, Apr. 2019.

[3] F. Kulwa, C. Li, X. Zhao, B. Cai, N. Xu, S. Qi, S. Chen, and Y. Teng, "A state-of-the-art survey for microorganism image segmentation methods and future potential," *IEEE Access*, vol. 7, pp. 100243–100269, 2019.

[4] C. Li, *Content-Based Microscopic Image Analysis*, vol. 39. Berlin, Germany: Logos Verlag Berlin GmbH, 2016.

[5] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," 2017, *arXiv:1712.04621*. [Online]. Available: http://arxiv.org/abs/1712.04621

[6] S. Liu, J. Zhang, Y. Chen, Y. Liu, Z. Qin, and T. Wan, "Pixel level data augmentation for semantic image segmentation using generative adversarial networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1902–1906.

[7] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019.

[8] X. Zhu, Y. Liu, Z. Qin, and J. Li, "Data augmentation in emotion classification using generative adversarial networks," 2017, *arXiv:1711.00648*. [Online]. Available: http://arxiv.org/abs/1711.00648

[9] Y. Zou, C. Li, K. Shirahama, T. Jiang, and M. Grzegorzek, "Environmental microorganism image retrieval using multiple colour channels fusion and particle swarm optimisation," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 2475–2479.

[10] Y. Zou, L. Chen, K. Shiriham, F. Schmidt, J. Tao, and M. Grzegorzek, *Content-Based Microscopic Image Retrieval of Environmental Microorganisms Using Multiple Colour Channels Fusion*. Cham, Switzerland: Springer, 2016.

[11] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.

[12] L. Deng, "The MNIST database of handwritten digit images for machine learning research [Best of the Web]," *IEEE Signal Process. Mag.*, vol. 29, no. 6, pp. 141–142, Nov. 2012.

[13] C. Li, K. Shirahama, and M. Grzegorzek, "Application of content-based image analysis to environmental microorganism classification," *Biocybernetics Biomed. Eng.*, vol. 35, no. 1, pp. 10–21, 2015.

[14] Q. You, J. Luo, H. Jin, and J. Yang, "Robust image sentiment analysis using progressively trained and domain transferred deep networks," in *Proc. AAAI*, 2015, pp. 1–9.

[15] G. Kang, X. Dong, L. Zheng, and Y. Yang, "PatchShuffle regularization," 2017, *arXiv:1707.07103*. [Online]. Available: http://arxiv.org/abs/1707.07103

[16] H. Inoue, "Data augmentation by pairing samples for images classification," 2018, *arXiv:1801.02929*. [Online]. Available: http://arxiv.org/abs/1801.02929

[17] A. Krizhevsky, V. Nair, and G. Hinton. (2009). *The CIFAR-10 Dataset*. [Online]. Available: http://www.cs.toronto.edu/~kriz/cifar.html

[18] T. Konno and M. Iwazume, "Icing on the cake: An easy and quick post-learnig method you can try after deep learning," 2018, *arXiv:1807.06540*. [Online]. Available: http://arxiv.org/abs/1807.06540

[19] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic minority over-sampling technique," *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, Jun. 2002.

[20] S.-M. Moosavi-Dezfooli, A. Fawzi, and P. Frossard, "DeepFool: A simple and accurate method to fool deep neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2016, pp. 2574–2582.

[21] J. Su, D. V. Vargas, and K. Sakurai, "One pixel attack for fooling deep neural networks," *IEEE Trans. Evol. Comput.*, vol. 23, no. 5, pp. 828–841, Oct. 2019.

[22] M. Zajac, K. Zolna, N. Rostamzadeh, and P. Pinheiro, "Adversarial framing for image and video classification," in *Proc. AAAI*, vol. 33, 2019, pp. 10077–10078.

[23] Y. Xue, Q. Zhou, J. Ye, L. Long, S. Antani, C. Cornwell, Z. Xue, and X. Huang, "Synthetic augmentation and feature-based filtering for improved cervical histopathology image classification," in *Proc. MICCAI*, 2019, pp. 387–396.

[24] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321–331, Dec. 2018.

[25] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, "GAN-based synthetic brain MR image generation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 734–738.

[26] H. Shin, N. A. Tenenholtz, J. K. Rogers, C. G. Schwarz, M. L. Senjem, J. L. Gunter, K. Andriole, and M. Michalski, "Medical image synthesis for data augmentation and anonymization using generative adversarial networks," in *Proc. MICCAI*, 2018, pp. 1–11.

[27] C. Han, K. Murao, T. Noguchi, Y. Kawata, F. Uchiyama, L. Rundo, H. Nakayama, and S. Satoh, "Learning more with less: Conditional PGGAN-based data augmentation for brain metastases detection using highly-rough annotation on MR images," in *Proc. 28th ACM Int. Conf. Inf. Knowl. Manage.*, Nov. 2019, pp. 119–127.

[28] C. Han, L. Rundo, R. Araki, Y. Nagano, Y. Furukawa, G. Mauri, H. Nakayama, and H. Hayashi, "Combining noise-to-image and image-to-image GANs: Brain MR image augmentation for tumor detection," *IEEE Access*, vol. 7, pp. 156966–156977, 2019.

[29] V. Sandfort, K. Yan, P. J. Pickhardt, and R. M. Summers, "Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks," *Sci. Rep.*, vol. 9, no. 1, pp. 1–9, Dec. 2019.

[30] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional GANs," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8798–8807.

[31] L. Taylor and G. Nitschke, "Improving deep learning with generic data augmentation," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2018, pp. 1542–1547.

[32] P. Pławiak, M. Abdar, J. Pławiak, V. Makarenkov, and U. R. Acharya, "DGHNL: A new deep genetic hierarchical network of learners for prediction of credit scoring," *Inf. Sci.*, vol. 516, pp. 401–418, Apr. 2020.

[33] P. Pławiak, M. Abdar, and U. Rajendra Acharya, "Application of new deep genetic cascade ensemble of SVM classifiers to predict the Australian credit scoring," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105740.

[34] P. Pławiak and U. R. Acharya, "Novel deep genetic ensemble of classifiers for arrhythmia detection using ECG signals," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11137–11161, 2019.

[35] M. Hammad, P. Pławiak, K. Wang, and U. R. Acharya, "ResNet-Attention model for human authentication using ECG signals," *Expert Syst.*, 2020, Art. no. e12547.

[36] T. Tuncer, F. Ertam, S. Dogan, E. Aydemir, and P. Pławiak, "Ensemble residual network-based gender and activity recognition method with signals," *J. Supercomput.*, vol. 76, no. 3, pp. 2119–2138, Mar. 2020.

[37] X. Yu, H. Wang, W.-C. Feng, H. Gong, and G. Cao, "GPU-based iterative medical CT image reconstructions," *J. Signal Process. Syst.*, vol. 91, nos. 3–4, pp. 321–338, Mar. 2019.

[38] X. Yu, H. Wang, W. Feng, H. Gong, and G. Cao, "An enhanced image reconstruction tool for computed tomography on GPUs," in *Proc. CF*, 2017, pp. 97–106.

[39] M. Hidayetoğlu, T. Biçer, S. G. de Gonzalo, B. Ren, D. Gürsoy, R. Kettimuthu, I. T. Foster, and W.-M.-W. Hwu, "MemXCT: Memory-centric X-ray CT reconstruction with massive parallelization," in *Proc. Int. Conf. High Perform. Comput., Netw., Storage Anal.*, Nov. 2019, p. 85.

[40] S. Kosov, K. Shirahama, C. Li, and M. Grzegorzek, "Environmental microorganism classification using conditional random fields and deep convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 248–261, May 2018.

[41] C. Li, K. Shirahama, and M. Grzegorzek, "Environmental microorganism classification using sparse coding and weakly supervised learning," in *Proc. 2nd Int. Workshop Environ. Multimedia Retr. (EMR)*, 2015, pp. 9–14.

[42] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Proc. NIPS*, 2014, pp. 2672–2680.

[43] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*. [Online]. Available: http://arxiv.org/abs/1511.06434

[44] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *Proc. Int. Conf. Mach. Learn.*, vol. 70, Aug. 2017, pp. 214–223.

[45] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville, "Improved training of wasserstein gans," in *Proc. NIPS*, I. Guyon, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017, pp. 5767–5777.

[46] Goldberger, Gordon, and Greenspan, "An efficient image similarity measure based on approximations of KL-divergence between two Gaussian mixtures," in *Proc. 9th IEEE Int. Conf. Comput. Vis.*, vol. 1, Oct. 2003, pp. 487–493.

[47] W. Hager, "Lipschitz continuity for constrained processes," *SIAM J. Control Optim.*, vol. 17, no. 3, pp. 321–338, 1979.

[48] F. Chollet. *Keras: Deep Learning Library for Theano and Tensorflow*. [Online]. Available: https://keras.io/k

[49] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, and, "TensorFlow: A system for large-scale machine learning," in *Proc. USENIX Symp. Operating Syst. Design Implement.*, 2016, pp. 265–283.

[50] NVIDIA. (2020). *NVIDIA GEFORCE RTX 2080Ti 11GB*. [Online]. Available: https://www.nvidia.cn/geforce/graphics-cards/rtx-2080-ti

[51] H. Salehinejad, S. Valaee, T. Dowdell, and J. Barfett, "Image augmentation using radial transform for training deep neural networks," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3016–3020.

[52] T. Jia, H. Chen, and J. Tang, "A research on generative adversarial network algorithm based on GPU parallel acceleration," in *Proc. Int. Conf. Image Video Process., Artif. Intell.*, Nov. 2019, pp. 397–404.

[53] T. Park, M.-Y. Liu, T.-C. Wang, and J.-Y. Zhu, "Semantic image synthesis with spatially-adaptive normalization," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 2337–2346.

HAO XU received the B.E. degree from Heilongjiang University, China, in 2017. He is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, MBIE College of Northeastern University, China. His research interests are microscopic image analysis, machine learning, and data augmentation.

H. Xu *et al.*: EF-GANs for Environmental Microorganism Image Augmentation With Limited Rotation-Invariant Training Data

IEEE *Access*

CHEN LI received the B.E. degree from the University of Science and Technology Beijing, China, in 2008, the M.Sc. degree from the Northeast Normal University, China, in 2011, and the Dr.-Ing. degree from the University of Siegen, Germany, in 2016. From 2016 to 2017, he worked as a Postdoctoral Researcher with the Johannes Gutenberg University of Mainz, Germany. He is currently working as an Associate Professor with Northeastern University, China. He is the Head of the Research Group for Microscopic Image and Medical Image Analysis, College of Medicine and Biological Information Engineering, Northeastern University. His research interests are microscopic image analysis, medical image analysis, machine learning, pattern recognition, machine vision, and multimedia retrieval. He is a reviewer for several journals and conferences, including *Pattern Recognition*, *Future Generation Computer Systems*, *Artificial Intelligence in Medicine*, *Chemometrics and Intelligent Laboratory Systems*, IEEE ACCESS, *Neurocomputing*, *Journal of X-Ray Science and Technology*, AAAI-20, and ITIB-2020.

MD MAMUNUR RAHAMAN received the B.Sc. degree from the BRAC University, Dhaka, Bangladesh, in 2017. He is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, Biomedical and Information Engineering School, Northeastern University. His research interests are microscopic image analysis, medical image analysis, machine learning, pattern recognition, and machine vision.

YUDONG YAO (Fellow, IEEE) received the B.Eng. and M.Eng. degrees in electrical engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 1982 and 1985, respectively, and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, in 1988. From 1987 to 1988, he was a Visiting Student with Carleton University, Ottawa, Canada. From 1989 to 2000, he was with Carleton University, Spar Aerospace Ltd., Montreal, Canada, and Qualcomm Inc., San Diego, CA, USA. Since 2000, he has been with the Stevens Institute of Technology, Hoboken, NJ, USA, where he is currently a Professor and the Chair of the Department of Electrical and Computer Engineering. He holds one Chinese patent and 13 U.S. patents. His research interests include wireless communications, machine learning and deep learning techniques, and healthcare and medical applications. For his contributions to wireless communications systems, he was elected as a Fellow of the National Academy of Inventors in 2015, and the Canadian Academy of Engineering in 2017. He has served as an Associate Editor for the IEEE COMMUNICATIONS LETTERS from 2000 to 2008, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2001 to 2006, and an Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS from 2001 to 2005.

ZIHAN LI was born in 1995. He received the B.E. degree from Northeastern University, China, in 2018. He is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, Northeastern University, China. His research interests are microorganism image retrieval and deep learning.

JINGHUA ZHANG was born in 1996. He received the B.E. degree from Hefei University, China, in 2018. He is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, Northeastern University, China. His research interests are microscopic image segmentation and deep learning.

FRANK KULWA was born in Tanzania. He received the B.E. degree from the Dar es Salaam Institute of Technology, Tanzania, in 2013. He is currently pursuing the master's degree with the Research Group for Microscopic Image and Medical Image Analysis, Northeastern University, China. Since 2017, he has been working as a Tutorial Assistant with the Dar es Salaam Institute of Technology. His research interests are microorganism image analysis and deep learning.

XIN ZHAO was born in 1982. He received the Ph.D. degree from the Harbin Institute of Technology, China, in 2011. He is currently an Associate Professor with the Environmental Engineering Institute, Northeastern University, China. He is also the Head of the Environmental Engineering Institute. His research interests are microbial molecular ecology and systems biology, water pollution control, and waste recycling.

SHOULIANG QI (Member, IEEE) received the Ph.D. degree from Shanghai Jiao Tong University, in 2007. He is currently an Associate Professor with the Sino-Dutch Biomedical and Information Engineering School, Northeastern University, China. He joined the GE Global Research Center, where he was responsible for designing innovative magnetic resonance imaging (MRI) system. From 2014 to 2015, he was a Visiting Scholar with the Eindhoven University of Technology and the Kempenhaeghe Epilepsy Center, The Netherlands. In recent years, he has been conducting productive studies in intelligent medical imaging computing and modeling, machine learning, brain networks, and brain models. He had published more than 80 papers in peer-reviewed journals and international conferences. He has won many academic awards, such as the Chinese Excellent Ph.D. Dissertation Nomination Award and the Award for Outstanding Achievement in Scientific Research from the Ministry of Education.
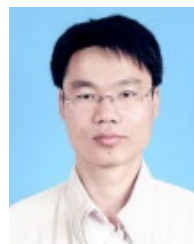
YUEYANG TENG worked with the PET-CT R&D Department, Shenyang Neusoft Medical System Co., Ltd., from 2005 to 2013. Since 2013, he has been working with Northeastern University, China. He is currently an Associate Professor with the School of Medicine and Bioinformatics, Northeastern University, and a Young Director of the Artificial Intelligence Branch of the Biomedical Engineering Society. His research areas include machine learning (especially deep learning) theory and application, neuroinformatics, and biomedical imaging.

• • •