Res-WGAN: Image Classification for Plant Small-scale Datasets

Mi Jiaqi
CAU
Hao Xia
Ph.D
Yang Si
Ph.D
Gao Wanlin
Ph.D
Li Minzan
Ph.D
Minjuan Wang (minjuan@cau.edu.cn)
China Agricultural University https://orcid.org/0000-0002-7520-1726
Zheng Lihua
Ph.D

Research

Keywords: Image classification, Plant small-scale dataset, Generative Adversarial 15 Network, Data enhancement

Posted Date: February 18th, 2020

DOI: https://doi.org/10.21203/rs.2.23790/v1

Res-WGAN: Image Classification for Plant Small-scale Datasets

Jiaqi Mi[1, 2], Xia Hao [2], Si Yang[1, 2], Wanlin Gao[2], Minzan Li[1, 2], Minjuan Wang[1, 2, 3]*, Lihua Zheng[1, 2]*

[1]Key Laboratory of Modern Precision Agriculture System Integration Research, China Agricultural University, Ministry of Education, Beijing 100083 China

[2]Key Laboratory of Agricultural Informatization Standardization, China Agricultural University, Ministry of Agriculture, Beijing 100083 China

[3]College of Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan 251100 China

*Corresponding author: Lihua Zheng and Minjuan Wang

Abstract

Background: The artificial identification of rare plants is always a challenging problem in plant taxonomy. Although the convolutional neural network (CNN) in the deep learning method can better realize the automatic classification of plant samples through training, the model accuracy is difficult to reach the human eye discrimination due to the quantitative limit of training samples. Thus, effective data enhancement is vital to improve the generalization ability and robustness of deep learning models, especially for plant small-scale data classification task. Different from traditional methods, the Generative adversarial network (GAN) mimics original data distribution and produces new samples with similar features which can help classifiers equip with extraordinary generalization ability. It has not been studied that data enhancement for plant samples' characteristics with GAN since sliced bread.

Result: In this study, we present a novel GAN model named as Residual Wasserstein GAN (Res-WGAN) for data enhancement. To further adapt to plant small-scale datasets, residual blocks were introduced into the classic WGAN-GP as the basic network unit.

These blocks enrich the presentation skills and sustained parameters unchanged simultaneously. Moreover, we enforce the idea from SRGAN to take content loss into a final function, which guarantes the similarity between generated samples and original samples in high-dimensional features. Benefiting from these improvements, the proposed Res-WGAN expanded original datasets efficiently. We test it on the ResNet and the experimental results show that new datasets combined transfer learning significantly promoted the accuracy of classification, especially at testing data. To illustrate the generalization of the model, more particular small datasets are applied for expansion and classification in this paper.

Conclusions: Our works report competitive accuracy results than other data enhancement methods, and user study confirms it's an ideal alternative strategy for small-scale plant datasets enhancement. Developing robust and effective small-scale plants classification method to replace expert testimony, is highly relevant for agricultural automation development.

Keywords: Image classification, Plant small-scale dataset, Generative Adversarial Network, Data enhancement

Background

Plant taxonomy is the science that finds, identifles, describes, classifies, and names plants. It's one of the main branches of taxonomy. The main process of the identity of an unknow plant by comparison with previously collected specimens or with the aid of books or identification manuals. This task requires a great deal of expertise, and classification becomes more difficult when rare plants are involved. Recently, image classification based on deep learning has become a research hotspot, achieving remarkable progress in innovation. Convolution Neural Network [1] (CNN) is the poster child which extracts high-dimensional features of images with multi-level convolution and activation layers to implement classification. Compared with traditional machine learning methods, it is more suitable for image processing. Several researchers have constantly tried to explore the best structure so that the accuracy of classification could reach or exceed the human level [2-5]. However, one of the major problems that needs immediate attention is the model training, which not only consumes lots of computing resources but also requires a huge dataset. Some of the classic datasets used for training have thousands of samples. For example, the CIFAR-10 [6] dataset has 50000 color images, and the number of standard ImageNet datasets is as high as 1.2 million. But many plant datasets cannot achieve such scale, especially for rare samples. Traditional data enhancement [7] methods, such as rotation or cropping, are difficult to produce new similar features. They are more beneficial for resisting noise rather than improving the classification. How to achieve accurate taxonomy on rare plant datasets is, thus, a crucial problem to be solved.

Aiming at small sample classification problem, a single source small sample image classification model, based on latent semantics for the problem of insufficient sample size was proposed by Liu [8]. in their experiment, the three semantic relationships of scene categories, images and objects were analyzed firstly. These include the semantic relationship of different scene labels, the objects contained in the scene, and the visual vocabulary. Secondly, the potential semantic relationship was represented with the calculated similarity, and the scenes in other categories such as the less-category data were searched. Thirdly, the samples satisfying the migration conditions are learned to expand the number of small sample data sets. However, this method has higher demand for existing samples and requires additional annotation information, which is difficult to achieve in real-life scenarios. In addition, the calculation of similarity between images is not reliable, because there are no perfect methods currently for measuring the distance between images. Both the Structural Similarity Algorithm (SSIM) and Mean Square Error (MSE) have their own limitations.

Transfer learning is another common and efficient way to perform on small sample datasets [9-11]. A pre-trained network is a saved network that has been trained previously on large datasets. If the original data set is large enough and versatile, the spatial hierarchy of the features of the pre-trained network learning will effectively serve as a general model of the visual world. Therefore, the quality of the original datasets is crucial. Nevertheless, there are many problems to implement transfer learning on certain specific scenarios, such as medical image or rare plant image classification. Because it’s difficult to find other datasets with similar characteristics in essence.

The generative adversarial network (GAN) [12], as an emerging generative algorithm, has made great progress in the fields of image generation [13-20] and scene transformation [21-24] and so on. GAN can imitate the distribution of images to generate new samples with similar characteristics. It is able to make full use of the original datasets to dig out profound information. In this paper, we propose a new data enhancement method based on GAN. Specifically, our work mainly focuses on the following aspects. Firstly, we take a variant of WGAN-GP [17] to acclimate plant small-scale images, which can avoid pattern collapse well as the WGAN-GP [17]. It is critical to expand the original datasets. Residual blocks are introduced into the network, which effectively enrich the presentation skills and sustain parameters unchanged simultaneously. Secondly, we borrowed the idea from SRGAN [25] to take content loss into the final function, which guarantees the similarity between generated samples and the original samples in high-dimensional features. Therefore, the whole function pushes the solution to the natural images manifold. Content loss can be obtained by calculating the high dimensional's L2 distance between newly generated images and the samples from original datasets. And we use a trained VGG [4] model to capture those features. Lastly, we tested the expanded dataset on ResNet, a current mainstream CNN used for classification. The purpose was to know whether using expanded data set with Res-WGAN would significantly improve the accuracy of classification, especially on the testing data. Further, the experiment verified that our model can provide good migratory ability on other small sample data sets.

Methods

Prior research proved that rich training features were the key to small sample classification. This work focuses on making full use of original plant datasets. The detailed architecture and optimization process of the proposed framework named Res-WGAN will be clarified in the following subsection.

Structure of Res-WGAN

In an attempt to leverage the success of GAN for image generation, the Res-WGAN for small-scale plant datasets enhancement was proposed to enhance version of the standard WGAN-GP. It describes a new architecture that integrates the residual blocks with backbone network, content loss, and other auxiliary tricks, such as Adam optimizer, LReLu, and Dropout. Through skip connections between different layers, the expression of Generator is strengthened. But meanwhile, these connections introduce neither extra parameter nor computational burden. And content loss that guarantees the similarity between generated samples and original samples in high-dimensional features is helpful to classification. More details about Res-WGAN is shown in Figure 1.

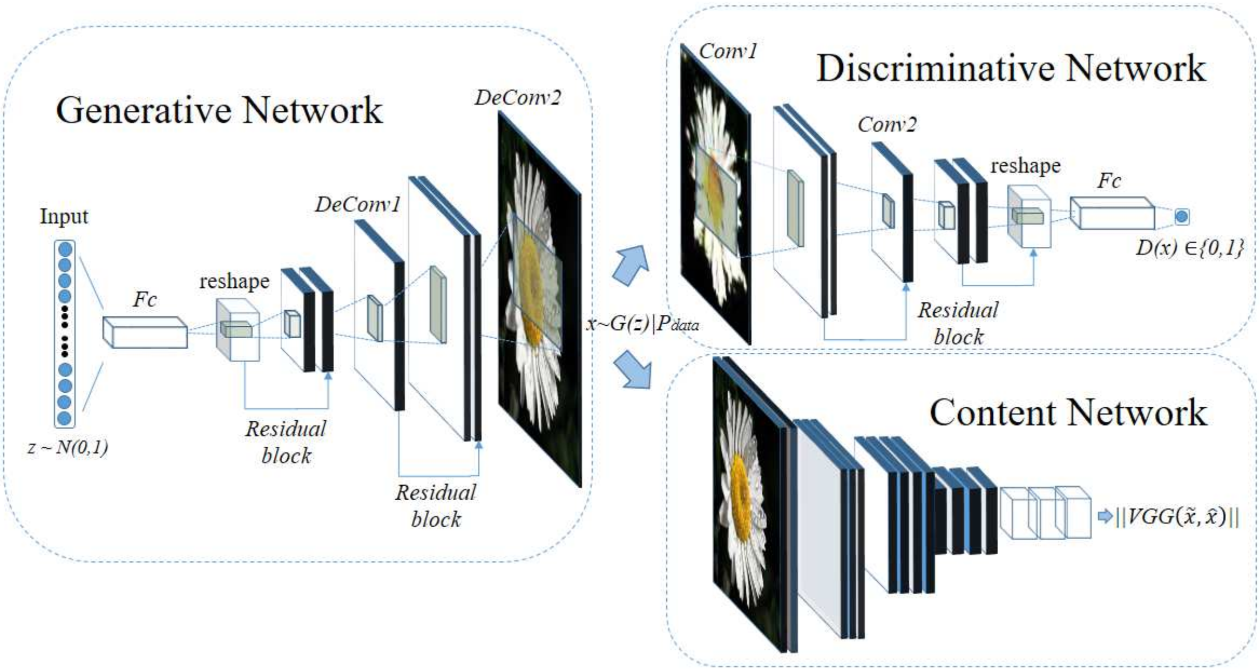


Fig.1 The overview of our proposed Res-WGAN. It mainly composed of generative network, discriminative network and content discriminator. Generative network contains one fully-connect layer, two residual blocks and two deconvolution layers, which are responsible for feature learning from random noise z to training sample $G(z)$. Discriminative network is made up with two convolution layers, one fully-connect layer and residual connections, and content discriminator is a

part of pre-trained VGG model. They are two parts of the whole discriminator, which distinguish whether the inputted image is derived from the original dataset P_r or generated images $G(z)$.

The architecture of Res-WGAN is a combination of three parts: generative network (shown in figure 1), discriminative network and content discriminator. The former is responsible for converting random noise z into images $G(z)$ that are similar to the sample x from the original dataset p_r visually. More specifically, z is a 100-dimensional random noise following a normal distribution, and it is projected to a small spatial extent convolutional representation with many feature maps. A series of fractionally-strided deconvolutions then convert this high-level representation into a 64×64-pixel image. The whole generator mainly contains two deconvolutional layers (i.e., *DeConv1, DeConv2*), two residual blocks (i.e., *Res-DeConv1-1, Res-Conv1-2, Res-DeConv2-1, Res-Conv2-2*), one fully-connected layer (i.e., *Fc*). We employed the basic architecture of ResNet and introduced residual blocks into the generator as the basic network unit. For example, the final output of the second residual block is equal to the sum of basic output of the *DeConv2* layer and the result of *Res-Conv2-2* layer. As shown in Figure 2, each block ends with a convolutional layer whose stride is set to 1×1. It's helpful to reduce the whole parameter count. More information about parameter settings are revealed in Table 1.

Table 1 Details of Res-WGAN layers

Generative Network						
Layer	Type	Input size	Patch size	Kernel sum	Stride	Parameter
Fc	Fully-connected	100	1×1	4×4×512	-	8192× (100+1)
Res-DeConv1-1	Deconvolution	4×4×512	5×5	256	2×2	(5×5+1) ×256
Res-Conv1-2-	Convolution	8×8×128	3×3	256	1×1	(3×3+1) ×256
DeConv1	Deconvolution	8×8×256	5×5	128	2×2	(5×5+1) ×128
Res-DeConv2-1	Deconvolution	16×16×128	5×5	64	2×2	(5×5+1) ×256

Res-Conv2-2	Convolution	16×16×64	3×3	64	1×1	(3×3+1) ×256
DeConv2	Deconvolution	32×32×128	5×5	3	2×2	(5×5+1) ×3
Output	-	64×64×3	-	-	-	-
Discriminator Network						
Conv1	Convolution	64×64×3	5×5	64	2×2	(5×5+1) ×64
Res-Conv1-1	Convolution	32×32×64	3×3	128	2×2	(3×3+1) ×128
Res-Conv1-2	Convolution	16×16×128	3×3	128	1×1	(3×3+1) ×128
Conv2	Convolution	16×16×128	5×5	256	2×2	(5×5+1) ×256
Res-Conv2-1	Convolution	8×8×256	3×3	512	2×2	(3×3+1) ×512
Res-Conv2-2	Convolution	4×4×512	3×3	512	1×1	(3×3+1) ×512
Fc	Fully-connected	4×4×512	1×1	1	-	1× (8192+1)
Output	-	1	-	-	-	-

The operating principle of Generative adversarial network is introduced in the paper GAN [13] of Goodfellow in detailly. In this paper, we hold the main idea of original GAN and enhanced each part to adapt to enhancement of plant small-scale datasets. The loss function of the generative network can be expressed as:

$$L_G = \mathbb{E}_{x \sim p_z}\left[D\left(G\left(x\right)\right)\right] - \mu \mathbb{E}_{\tilde{x} \sim p_r, \hat{x} \sim p_z}\left[\left\| VGG\left(\tilde{x}\right) - VGG\left(G\left(\hat{x}\right)\right) \right\|_2\right] \tag{1}$$

Where L_G represents the whole loss of the generative network that consists of two parts: the discriminant loss and the content loss. The former refers to a loss obtained by inputting the generated plant samples into the discriminator. Specifically, p_z represents random noise which obeys the normal distribution. It's sent into the generator *G* and gets the pictorial result $G(x)$. Then the newly generated images are taken as the input of discriminator *D* to distinguish between real and fake, i.e., $D\big(G(x)\big)$. $\mathbb{E}$ is an expectation, which means the average value is calculated multiple times during the actual operation. The latter refers to inputting the generated images by *G* and samples from original plant datasets p_r into the pre-trained VGG network at the same time to obtain the L2 distance between them. Content loss can guarantee consistency between new samples and raw samples in high dimensional features that is vital for producing.

In the experiment, we randomly took eight images from dataset p_r into the VGG model and got the average output of the thirteenth layers as content feature per iteration. In order to meet the input size, we enlarge every image from 64×64 to 224×224 artificially. μ represents the weight coefficient, and we set it to 0.001 normally.

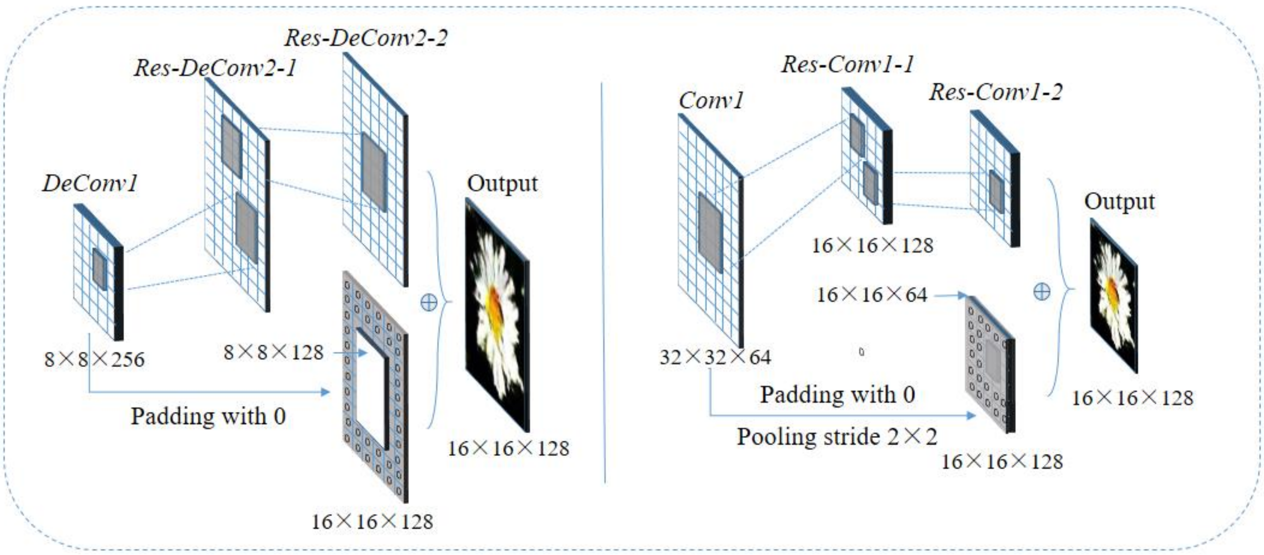


Fig. 2 The structure of residual blocks**.** In figure 2, the left part is the second residual block in generative network. It contains two deconvolution layers (i.e., *Deconv1*, *Res-Deconv2-1*) and one convolution layer (i.e., *Res-Deconv2-2*). The final output of this block equals to the basic result of *Res-Conv2-2* layer combined with the output of *Deconv1* layer. In order to meet the requirements of size, we took zero as padding. Similarly, the first residual block in discriminative network is shown on the right. There are three convolution layers (i.e., *Conv1, Res-Conv1-1, Res-Conv1-2*) in total.

From figure 1, the other part of Res-GAN is the discriminator. Indeed, the whole discriminator consist of a discriminative network and a content discriminator. Discriminative network is responsible for distinguishing whether the input comes from the ground truth or the generative network. That is a two-class problem in a sense. It mainly contains two convolutional layers (i.e., *Conv1, Conv2*), two residual blocks (i.e., *Res-Conv1-1, Res-Conv1-2, Res-Conv2-1, Res-Conv2-2*), one fully-connected layer (i.e., *Fc*). More parameter information can be obtained from Tabel 1. Similarly, the residual block was brought in to enhance the ability of discriminative network. For instance, the final output of the first residual block is equal to the sum of basic output of the *Conv1* layer and the result of *Res-Conv1-2* layer. More specific connection refers to Figure 2. The content discriminator is designed to calculate high-dimensional features' difference between two images, described in SRGAN [25] with detail. Specifically, we took the VGG-16 as content discriminator, and performed prior training on dataset ImageNet [28]. Due to the abundance of species in ImageNet, the classification model is equipped with a good generalization ability. The final loss function of the discriminative network is shown in Equation 3.

$$L_D = -\mathbb{E}_{x\sim P_r}\left[D(x)\right] + \mathbb{E}_{x\sim P_z}\left[D(G(x))\right] + \lambda\mathbb{E}_{x\sim P_{\hat{x}}}\left[\left\|\nabla_x D(x)\right\|_p - 1\right]^2 \tag{2}$$

where L_D is the final loss of the discriminator, which includes the discriminative loss and gradient penalty term. The former refers to the sum of the loss values obtained by inputting the generated samples and the original samples into the discriminative network. p_r represents the original plant dataset, and p_z is the random noise following a normal distribution. $G(x)$ is the image generated by the generative network G from p_z. $D(x)$ means gaining the loss value by inputting p_r or p_z into the discriminative network D. The gradient penalty term is derived from WGAN-GP. The purpose of the setting is to establish a connection between the training gradient of the discriminative network and the hyperparameter K(usually set to 1), so that the function represented by the network satisfies the condition - Lipschitz continuity. λ represents the weight coefficient, and we set it to 10.0 normally. In conclusion, the final function of Res-WGAN can be expressed as:

$$\arg\min_G \max_D \mathbb{E}_{x\sim P_r}\left[D(x)\right]-\mathbb{E}_{x\sim P_z}\left[D\big(G(x)\big)\right]-\boldsymbol{\lambda}\mathbb{E}_{x\sim P_{\hat{x}}}\left[\left\|\nabla_x D(x)\right\|_p-1\right]^2+ \boldsymbol{\mu}\mathbb{E}_{x\sim P_r,x\sim P_z}\left[\left\|VGG(\tilde{x})-VGG\big(G(\hat{x})\big)\right\|_2\right] \tag{3}$$

Here, the ultimate objective is to minimize the generative loss whilst maximizing the discriminative loss. The discriminative loss here refers to the result of the overall discriminative network, as the generative loss. For the independence of training, we first look at the $\max D$ part. The goal of discriminative network D is to distinguish between the true to fake inputs correctly. Here 0 means fake and 1 means true. If the sample x from the original dataset p_r is sent to D, then $D(x)$ is expected to approach 1. Similarly, in the second term, x from random noise P_z is converted into image $G(x)$ by the generative network G, so that $D\big(G(x)\big)$ is expected to approach zero. Thirdly, as the gradient penalty $||\nabla_x D(x)||_p$, is asked to be closer to 1, the whole penalty term $||\nabla_x D(x)||_p - 1$ comes near to 0. In short, this part is expected to make the overall loss larger i.e. $\max D$. And $\min G$ is similar to $\max D$, in this case, D is kept unchanged. Only the second and fourth items are useful for the whole function. G is trained to spoof D to get $D\big(G(x)\big)$ close to 1. The fourth item represents the formula of content loss. It's required to be as small as possible. So, this part is expected to make the whole minimum i.e. $\min G$. The details of training are shown in Table 2.

Table 2 The training process of Res-WGAN

Algorithm 1 Res-WGAN with gradient penalty. We used default values of $\lambda = 10, \mu = 0.001,\ n_{critic} = 5,\ \alpha = 0.0001,\ \beta_1 = 0.9,\ \beta_2 = 0.99$.

Require: The gradient penalty coefficient λ, the number of critic iterations per generator iteration n_{critic}, the batch size m, Adam hyperparameters α, β_1, β_2.

Require: Initial critic parameters w_0, initial generator parameters θ_0.

while θ has not converged **do**

  for $t = 1, \dots, n_{critic}$ **do**

    for $i = 1, \dots, m$ **do**

      Sample real data $x \sim P_r$, latent variable $z \sim P_z$, a random number $\epsilon \sim U[0,1]$.

      $\tilde{x} \leftarrow G_\theta(z)$

      $\hat{x} \leftarrow \epsilon x + (1-\epsilon)\tilde{x}$

      $L_D^{(i)} \leftarrow D_w(\tilde{x}) - D_w(x) + \lambda(\|\nabla_{\tilde{x}} D_w(\hat{x})\|_2 - 1)^2$

    end for

    $w \leftarrow Adam\left(\nabla_w \frac{1}{m}\sum_{i=1}^{m} L_D^{(i)}, w, \alpha, \beta_1, \beta_2\right)$

  end for

  for $j = 1, \dots, m$ **do**

    Sample real data $x \sim P_r$, latent variable $z \sim P_z$, a trained well network VGG

    $L_G^{(i)} \leftarrow D_w(G_\theta(z)) + \mu\left\|VGG(\tilde{x}) - VGG(G(\hat{x}))\right\|_2$

  end for

  $\theta \leftarrow Adam\left(\nabla_w \frac{1}{m}\sum_{i=1}^{m} -L_G^{(i)}, \theta, \alpha, \beta_1, \beta_2\right)$

end while

Results

Classification on small-scale datasets

It's a proverbial phenomenon that classification accuracy tends to decrease as dataset shrinking, but the baseline where underfitting happens waits to be proved. In this part, we gradually reduced the number of training samples and visualized the accuracy change on training and testing data. The ResNet has been chosen as the classifier. To ensure the generality of results, lots of experiments has been done on benchmark datasets. The result on Flowers and Beans are demonstrated in Figure.2 as common examples. The Flowers is a classic plant dataset, which contains five varieties of flowers, such as dandelion, sunflower, tulip and so on. Each species has about 1000 pictures. And the Beans contains 8000 images of soybean leaves under 10 different stresses in total.

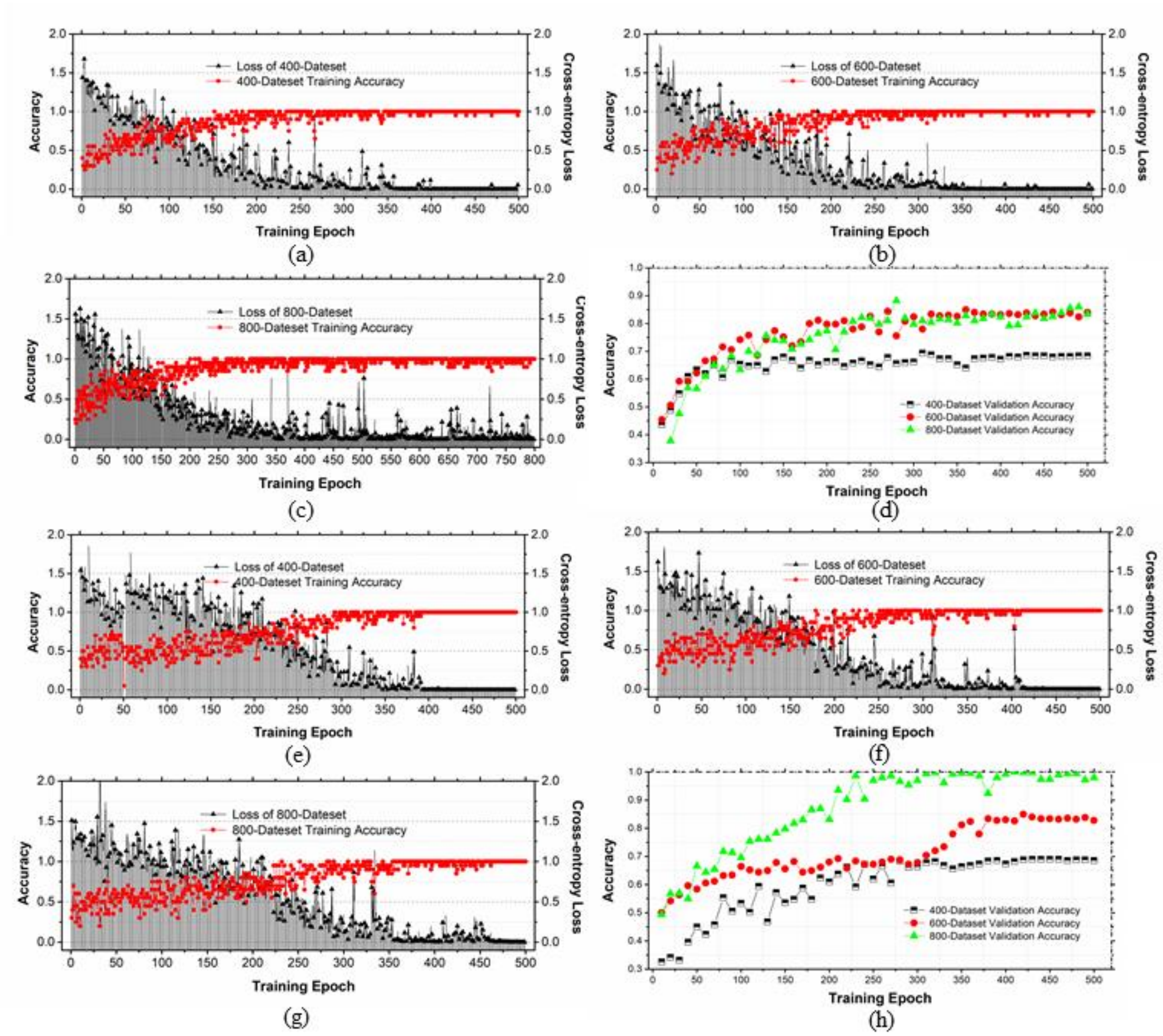


Fig.3 The trend of training and validation accuracies varying with different size datasets.On the dataset Flowers, (a), (b) and (c) show the training accuracy when the number of a single type of dataset is 400, 600 and 800 respectively, whereas their validation accuracy is shown in (d). Similarly, (e), (f), (g) and (h) are the result on dataset Beans.

In Figure 3, as the size of the training data set decreases, the accuracy of the classification model ResNet on the training data does not change significantly. It's obvious that their accuracy reaches nearly one and the loss reaches zero. It is significant from the training point of view, because something has been learned by the ResNet. Unfortunately, the accuracy of testing data comes up with a bigger change due to the decrease in the training data. As the size of the training data set decreases, the accuracy of ResNet on the testing data also decreases. Thus, over-fitting phenomenon caused obviously in the case a single type of data set is 400. Specifically, it showed high accuracy on training set, but low on test set. Many auxiliary tricks, like Dropout and LReLu, have been used in training progress to prevent the problem. But it's indeed difficult for the classification model to learn effective image features from actually small data sets.

Evaluations

In this part, Res-WGAN is adequately compared with other models or methods from generative ability and accuracy improvement. From above charts, prediction on testing data declines heavily when the sample size is set to 400 each class, so a new study is carried out on this basis.

Structure comparison

Time complexity and space complexity are important indexes for evaluating algorithm. To prove our method hasn't brought more space consumption and easy to train, we compare it with other state-of-the-art GANs as shown below in Table 3.

Table 3 Parameters, FLOPs, Training Epoch, Training Time of models during training phase

Model	Parameter	FLOPs	Training Epoch	Training Time (h)	IS
Res-WGAN	25255621	304907093851	62	**3.5**	30.6
DCGAN	35701637	998396935281	80	5.0	18.5

WGAN	28328837	249636095476	67	**3.5**	21.2
WGAN-GP	28328837	395326772642	54	4.0	25.6
SAGAN	59257754	714406331436	**45**	4.5	**31.5**

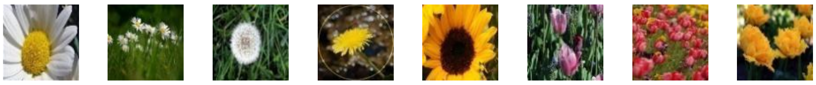
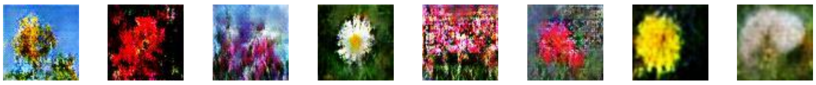
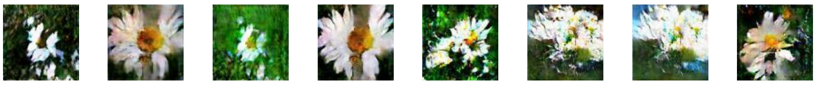
Table 3 quantifies the comparisons between Res-WGAN and state-of-the-arts, which can be regarded as strong proof for the effectiveness of our model. Although parameters are declined into the lowest level in Res-WGAN compared with DCGAN, WGAN-GP, SAGAN, it also needs more calculation amount and time for training in terms of FLOPs and Training Time. Inception Score (IS) is a common measurement index of quality and diversity of generated image. In this paper, 400 generated images each class have been used to calculate IS. Our model obtains the second level 30.6 just surpassed by SAGAN's 31.5. Moreover, SAGAN is famous for generating high quality images. Because of the plant small samples dataset's restriction, all the models get a lower score than training on larger dataset in other papers.

Generative ability

In this part, we continue to confirm the superiority of Res-WGAN on Flowers. Obviously, visual images are strong proof of generative models. In addition to image sharpness and distortion, the outputs diversity is another indispensable index to evaluate GAN. The outputs of DCGAN, WGAN, SAGAN and Res-WGAN with sufficient iteration are shown in Table.4. During the training process, the whole dataset was fed into the model with compressed size 64×64. And the outputs of generator were the same size.

Table 4 Outputs of different generative models training on Flowers

Model	Result

Raw Date	
DCGAN	
WGAN-GP	
SAGAN	
Res-WGAN	

As Table 4 depicts, the images generated by the original WGAN and DCGAN are obviously blurry, with disordered quality, and the texture details are also missing. As the expressive power of these complex networks is limited due to small sample training datasets, the network is hard to learn better even with more training. This point can be confirmed by increasing the size of the training data. Residual blocks structure is introduced into the Res-WGAN model proposed in this paper, which reduces the network levels and enhances the expressive ability. Thus, it generates better images on a small sample dataset as shown in Fig. 4. Moreover, the result shows that the images generated by Res-WGAN are clearer than other networks, and the details of texture are more orderly as well. It's easier to classify these images visually. On the other hand, the pattern collapse occurred in SAGAN and DCGAN in terms of the types of images generated. The details about it were explained in the raw WGAN paper [16]. In fact, this is a flaw with the model itself. However, it is destructive for the generative model because it is necessary to get more images that are different from the original small samples. The improved Res-WGAN avoids this situation. It accomplished the expanding of the original data by controlling the input values of the model.

Classification improvement

In this part, the expanded dataset was used for training of ResNet, and we get the new classification accuracy. To fully verify the enhancement effect of the expanded dataset with Res-WGAN, other general enhancement methods like rotation, random erasing, cropping and transfer learning have been added into the experiment (Table 5).

Table 5 Accuracy of training and validation on different size dataset with different models

Dataset size	400 (train)	400 (test)	600 (test)	800 (test)
Raw data	0.998±0.001	0.682±0.002	0.835±0.005	0.860±0.003
Transfer learning	0.998±0.001	**0.829±0.020**	/	/
General Methods	/	/	0.680±0.020	0.695±0.005
DCGAN	/	/	0.697±0.004	0.694±0.002
WGAN	/	/	0.701±0.015	0.706±0.004
WGAN-GP	/	/	0.723±0.005	0.718±0.002
SAGAN	/	/	0.694±0.007	0.702±0.010
Res-WGAN	/	/	0.746±0.005	**0.751±0.002**
Transfer learning + Res-WGAN	/	0.829±0.020	**0.832±0.010**	**0.837±0.005**

In Table 5, it's clear that the training accuracy of the ResNet already approaches one when the sample size each class is set to 400. So, the duplicate results are removed when the size of dataset is 600 or 800. Because of the bottleneck ResNet has reached in training, it is difficult to take one step further. Moreover, there are some poor-quality images in the dataset that are hard to classify. Improving the model's accuracy on testing data is the key point in this paper. This is because the classification model usually causes over-fitting on small-scale training data (as shown in Table 2 (d) and (h)), although some countermeasures like Dropout or Leaky ReLU have been taken for that. Specifically, Specifically, overfitting is characterized by excellent model performance on training data, but poor on test data. However, it is negligible in terms of the traditional data enhancement for accuracy on testing data compared with other methods. When the size of per category in original dataset is expanded from 400 to 600 or 800 with general methods, the accuracy on ResNet just improves by 1.3%. The main reason may be traditional methods can't make more extra features from the raw data. Although they promote the anti-interference ability of the model, the traditional methods do not contribute much faced with new samples. In addition, the result also shows that the combination of transfer learning and the Res-WGAN method promote the validation accuracy effectively. The difference between them is that the transfer learning method gets more features by pre-training on other datasets. And those new features help us classify on testing sets. When comparing the single method, transfer learning gets the highest validation accuracy (82.9%). Corresponding to the transfer learning, the purpose of the proposed method is to fundamentally learn the high-dimensional features from various images types and generate similar features, which in turn helps the classification model to learn. In the experiment, a one-to-one training mode is taken to produce new samples. In brief, the same kinds of flowers are generated one by one. The size of those images is fixed at 64×64, which is smaller than the size of images in original dataset but consistent with the input of the discriminative network. Indeed, it is difficult to output larger images because of the restriction of the small-scale samples. Finally, the combined dataset (generated images and original images) was used for the training of the ResNet and validation accuracy can reach 75.1%, which reached the third-highest accuracy in augmented datasets (shown in table 5). The results confirmed that our proposed Res-WGAN model can effectively learn features and enhance the generalization ability of the classification model.

Meanwhile, inspired by transfer learning, we use pre-trained model as base model and trained it on the extended dataset with Res-WGAN. It's so exciting that the validation accuracy (83.7%) exceeds all results for the same period. However, that result is only in line with the real training dataset size is equal to 600. There is still a gap from the best results. Compared with DCGAN and SAGAN, WGAN series performs better on small-scale plant datasets. That's maybe related to the model's generative capacity itself, and dig deep relationships between them is our future work.

Model migration

There are various forms of scenes in botanical research where our model can play an important role. In this part, the Res-WGAN model was used for other small datasets to verify its generalization ability. The dataset includes three kinds of green plants, including *Arabidopsis*, *Beans*, and Romaine Lettuce. Each class contains four hundred image samples. It's necessary to classify them accurately in actual experimental environment. According to the result of former studying, we expanded the original dataset from four hundred images each class to eight hundred with Res-WGAN. Then they were inputted to ResNet to test the accuracy. Table 6 shows the enhanced dataset obtained by Res-WGAN.

Table 6 Generated images of Res-WGAN training on small sample dataset

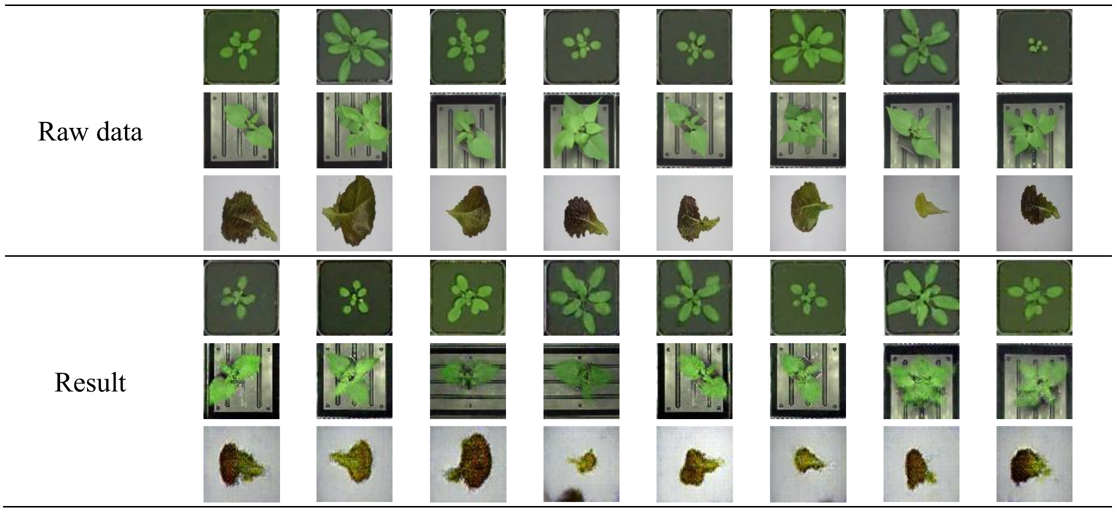


Table 7 Training accuracy, validation accuracy on different size dataset of models

Dataset size	400 (train)	400 (test)	600 (test)	800 (test)
Raw data	0.999±0.001	0.693±0.020	0.812±0.010	0.820±0.003
Transfer learning	0.999±0.001	**0.784±0.005**	/	/
General Methods	/	/	0.699±0.005	0.705±0.005
Res-WGAN	/	/	0.756±0.003	**0.759±0.002**
Transfer learning + Res-WGAN	/	0.784±0.005	0.798±0.002	**0.802±0.004**

The combination of transfer learning and Res-WGAN gets the best result than other state-of-the-arts (shown in Table 7). This is consistent with the results in section 4.2.3. It's just 1.8% less than the real dataset with 800 images in each class. It confirms that new samples generated by Res-WGAN enhance the generalization ability of the classification model once again. Undoubtedly, the proposed method in this paper is a favorable alternative to the original data set when the real size of the dataset is limited.

Discussion

Original training dataset is a constraints in the automated classification of rare plant samples using CNN. At the beginning of experiments, we explore the impact of training samples at different levels on the classification model. Two datasets of "Flowers" and "Beans" were selected to reflect objectivity. Forthermore, we have selected three orders of magnitude of 400, 600, and 800 with significant changes in the accuracy. They respectively identify the current number of samples of each class participating in classification. It can be seen (Figure 3 (d) and (h)) that the classification model has already experienced a serious overfitting phenomenon when the number is set to 400. If we continue to reduce the number, this phenomenon will become more serious and the research is of little significance. Therefore, the basic size of datasets is set to 400, which is convenient for comparing the effects of different enhancement methods on the classification in the next experiment. Unfortunately, the number of rare plant images can sometimes not reach the magnitude of 400. But as an exploratory method, our work solved the classification problem within this scale well. Future research will look at the classification on smaller plant samples.

At the main part of experiments, we have proved the superiority of proposed method in multiple angles. As far as the model structure, the model expressibility is enhanced by reducing the network layer and introducing residual connections. Meanwhile, parameters number is reduced. Therefore, the training of the model has reduced its dependence on the data, so it is very friendly to rare plant data. But the training time will increase because the residual structure introduces more calculations. In terms of the quality of new generated samples, Tab 4 makes a good comparison. From a visual aspect, the generated data obtained by Res-WGAN has clearer texture features, and it can better avoid the occurrence of model collapse. That's mainly attribute to the loss function, which was described in more detail before. In terms of the improvement in classification tasks, we compared this method with different enhancement methods. As for a single method, transfer learning is the most efficient method. By using pre-trained parameters on other datasets, the accuracy on the testing data can be improved from 0.682 to 0.829, which is close to the original sample size of 600 (0.835). Our method comes in second (0.751). From the results, it proved that the Res-WGAN indeed improves the accuracy of the classification model. At the same time, the content loss guarantes the similarity between generated samples and original samples in high-dimensional features. It's very helpful to the final classification. In the table, the Res-WGAN can reach the accuracy 0.746, which is 0.04 more than the best WGAN-GP in the classic GAN and 0.06 more than the traditional data enhancement method, when the size of training data is expanded to 600.

Nevertheless, the data enhancement use GAN alone for is unsatisfactory. Then we combine two state-of-the-art methods and obtain the result 0.832 that is almost equal to the result when original sample is 600. However, when the amount of training data is expanded to 800, the accuracy has not improved significantly. It indicates that this type of enhancement method has reached the bottleneck. More new image samples will not help classification, but introduce noise. That is already reflected in DCGAN and WGAN-GP. Finally, We verified our model on other datasets to prove the robustness of our new model. The experimental results fully meet our expectations as well.

There are still some shortcomings of Res-WGAN, although it performed well in image generation. We tried to use the model to generate larger scale sample data but ended with failure. The newly generated samples didn't form complete texture features but a lot of noises. We suspect that the small sample training data restricts the express ability of the model. Nevertheless, large-scale image samples cover more pixels, and it becomes more difficult to coordinate the relationship between pixels. This proves that there are certain constraints on the application of our model. If the amount of original sample data is too small, the model's generating ability will decreases significantly. Improving the quality of images from GAN based on small-scale samples is of major importance in our next work.

Conclusion

Due to the inaccessibility of plant large-scale datasets in real experiments, special techniques are required to investigate the precise classification. This paper proposed an effective data enhancement method based on the dedicated generative adversarial network. The main contributions and shortcomings are summarized as follows:

We brought up a novel GAN structure for special training datasets. Using Earth Mover's Distance (EMD) as the measure of the distance between images, and holding the original WGAN-GP structure as our backbone network. Residual blocks were introduced to enhance the model's express ability. These structures didn't bring additional parameters, but increased the amount of computation. Furthermore, we added the content loss into the final loss function that ensured the similarity between generated samples and original samples in high-dimensional features. This consistency was very helpful for subsequent classification. Experiments showed that, compared with other methods, the enhanced dataset with transfer learning improved the classification accuracy especially on testing data. The enhancement effect approximately equally to the original dataset with 600 images each class. At the same time, the Res-WGAN model obtained similar results when applied to other plant small-scale datasets. It confirms that our model has a strong generalization ability and it's an ideal alternative strategy for small-scale plant datasets enhancement.

Supplementary information

Abbreviations

CNN: convolutional neural network; GAN: Generative adversarial network; Res-WGAN: Residual Wasserstein GAN; SSIM: Structural Similarity Algorithm; MSE: Mean Square Error; EMD: Earth Mover's Distance

Acknowledgements

This work was supported by the National Key Research and Development Program (Grant No. 2016YFD0200600-2016YFD0200602) and the Key Research and Development Project of Shandong Province (Grant No. 2019GNC106091). All of the mentioned support is gratefully acknowledged. In addition, thanks for all the help of the teachers and students of the related universities.

Authors' contributions

JM and MW drafted and revised the manuscript. JM, WG, ML and LZ conceived of the study, and reviewed the manuscript. SY, XH and JM participated in the sample collection, experimental data acquisition. JM and MW analyzed and interpreted the results. All authors read and approved the fnal manuscript.

Funding

This work was supported by the National Key Research and Development Program (Grant No. 2016YFD0200600-2016YFD0200602) and the Key Research and Development Project of Shandong Province (Grant No. 2019GNC106091).

Availability of data and materials

All data generated or analysed during this study are included in this published article and in Additional file 1.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Key Laboratory of Modern Precision Agriculture System Integration Research, China Agricultural University, Ministry of Education, Beijing 100083 China

[2]Key Laboratory of Agricultural Informatization Standardization, China Agricultural University, Ministry of Agriculture, Beijing 100083 China

[3]College of Information Science and Engineering, Shandong Agriculture and Engineering University, Jinan 251100 China

*Corresponding author: Lihua Zheng and Minjuan Wang

References

Huang, K.Q., Ren, W.Q., Tan, T.N. A Review on Image Object Classification and Detection. Chinese Journal of Computers. 2014;36(6):1-18.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998;86(11):2278-2324.

Krizhevsky, A., Sutskever, I., Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in neural information processing systems, 2012;1097-1105.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., & Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition. 2015;1-9.

Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

He, K., Zhang, X., Ren, S., & Sun, J. Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition. 2016;770-778.

Krizhevsky, A., & Hinton, G. Learning multiple layers of features from tiny images. Technical report. 2009;7.

Chen, W.B., Guan, Z.X. Data augmentation method based on conditional generative adversarial net model. Jouranl of Computer Applications. 2018;38(11):3305-3311.

Liu, C.W. Small Sample Image Classification Based on Semantic Computation[D]. Doctor Thesis of Chongqing University, Chongqing, China. 2017.

Song, G.H. Image Annotation Method based on Transfer Learning and Deep Convolutional Feature[D]. Doctor Thesis of Zhejiang University, Hangzhou, China. 2016.

Wu G.Q. Research and Application of Transfer Learning in Image Classification. Master Thesis of Anhui University, Anhui, China. 2017.

Wang, H., Nie, F., Huang, H., & Ding, C. Dyadic transfer learning for cross-domain image classification. International Conference on Computer Vision. IEEE, 2011;551-556.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. Generative adversarial nets. Advances in neural information processing systems. 2014;2672-2680.

Radford, A., Metz, L., & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint, arXiv:1511.06434, 2015.

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. Image-to-image translation with conditional adversarial network. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;1125-1134.

Mao, X., Li, Q., Xie, H., Lau, R. Y., Wang, Z., & Paul Smolley, S. Least squares generative adversarial networks. Proceedings of the IEEE International Conference on Computer Vision. 2017;2794-2802.

Arjovsky, M., Chintala, S., Bottou, L. Wasserstein GAN. arXiv preprint, arXiv: 1701.07875, 2017.

Gulrajani, I., Ahmed, F., Arjovsky, M., et al. Improved training of wasserstein gans. Advances in neural information processing systems, 2017;5767-5777.

Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. Spectral normalization for generative adversarial networks. arXiv preprint. arXiv:1802.05957, 2018.

Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. Self-attention generative adversarial networks. arXiv preprint. arXiv:1805.08318, 2018.

Brock, A., Donahue, J., & Simonyan, K. Large scale gan training for high fidelity natural image synthesis. arXiv preprint. arXiv:1809.11096, 2018.

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision. 2017;2223-2232.

Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;1125-1134.

Miyato, T., & Koyama, M. cGANs with projection discriminator. *arXiv preprint*. arXiv:1802.05637, 2018.

Mao, Q., Lee, H. Y., Tseng, H. Y., Ma, S., & Yang, M. H. Mode seeking generative adversarial networks for diverse image synthesis. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019;1429-1437.

Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. Photo-realistic single image super-resolution using a generative adversarial network. Proceedings of the IEEE conference on computer vision and pattern recognition. 2017;4681-4690.

Valerio Giuffrida, M., Scharr, H., & Tsaftaris, S. A. ARIGAN: synthetic Arabidopsis plants using generative adversarial network. Proceedings of the IEEE International Conference on Computer Vision. 2017;2064-2071.

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. 2009;248-255.

Yang, Z. W., Zeng, S.Y., & Yang, Y. F. Face verification based on information fusion binary tree convolution neural network. Computer Applications. 2017;37(S2):155-159.

Zhao, X. X., Qian, S. S., & Liu, X. G. Image Identification Method for High-Strength Bolt Missing on Railway Bridge Based on Convolution Neural Network. China Railway Science.2018;39(4):56-62.

Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. Random erasing data augmentation. *arXiv preprint*.arXiv:1708.04896, 2017.

Yang, Y. N., Qi, L. H., Wang, H., & Su, L. P. (2019). Research on Generation Technology of Small Sample Data Based on Generative Adversarial Network. Electric power construction. 2019;71-77.

Mo, Z., Gai, Y. R., & Fang, G. L. Credit card fraud classification based on GAN-AdaBoost-DT imbalanced classification algorithm. Computer Applications. 2018;0-0.

Yi, K. Image recognition based on depth network ACGAN. Image & Multimedia Technology. 2018;(18):50.

Zhao, C. H., & Qi, B. Hyperspectral image classification based on Monte Carlo feature reduction method. Infrared Millim. Waves. 2013;62-67.

Shu, X., Yu, H. M., Zheng, W. W., Xie, Y., Hu, H. J., & Tang, H. M. Classifier-designing Algorithm on a Small Dataset Based onMargin Fisher Criterion and Transfer Learning. Acta automatica sinica. 2016;42(9):1313-1321.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.rar