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

Mi Jiaqi
China Agricultural University
Hao xia
china agricultural university
Yang Si
china agricultural university
Yang Xiande
Shandong agriculture and Engineering University
Gao Wanlin
china agricultural university
Li Minzan
china agricultural university
Zheng Lihua
China Agricultural University
Minjuan Wang (minjuan@cau.edu.cn)
China Agricultural University https://orcid.org/0000-0002-7520-1726

Research

Keywords: Image classification, Plant small-scale dataset, GAN, Data enhancement

Posted Date: June 2nd, 2020

DOI: https://doi.org/10.21203/rs.3.rs-31967/v1

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

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

[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 and Rural Affairs, Beijing 100083 China

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

*Corresponding author: Minjuan Wang and Lihua Zheng

Abstract

Background: Artificial identification of rare plants is an important yet challenging problem in plant taxonomy. Although deep learning-based method can accurately predict rare plant category from training samples, accuracy requirements of only few experts are satisfied due to the small sample size. Thus, effective data augmentation is vital to improve the expressive power and robustness of deep learning models, especially for plant small-sample classification tasks. Different from traditional methods, a generative adversarial network (GAN) can mimic the distribution of primary data and produce similar but nonidentical samples. Data augmentation for automated classification of rare plant samples with GAN has not been studied for a long time.

Result: In this study, we present a novel GAN model, referred to as residual Wasserstein GAN (Res-WGAN), for extending rare plant dataset efficiently. With a premise of minimized size of training parameters, residual block serves as elementary building block to increase the model's expression ability. A loss function based on Wasserstein distance with gradient penalty is used to ensure diversity of output. Moreover, we draw on the super-resolution GAN (SRGAN) and consider perceptual loss into the function. Perceptual loss guarantees similarity between generated samples and original samples in high-dimensional features. Benefiting from these improvements, the expanded dataset improves the classification accuracy on the residual neural network (ResNet), a classic convolutional neural network, and restrains the overfitting

phenomenon effectively by using transfer learning. More similar experiments are presented in this paper to prove the practicality of the model.

Conclusions: Our proposed method produces better diversity and sharpness of generated samples and more accurate classification results than other data augmentation methods. In addition, a user study confirms that it is an ideal alternative strategy for small-scale rare plant identification. Developing a robust and effective small-scale plant classification method to replace expert testimony is highly relevant for agricultural automation development.

Keywords: Image classification, Plant small-scale dataset, GAN, Data enhancement

Background

Plant taxonomy, one of the main branches of taxonomy, is a science that finds, identifies, describes, classifies, and names plants [1]. The main process of identifying an unknown plant is to compare it with previously collected specimens or with the aid of books or identification manuals. This task requires a great deal of expertise; thus, classification becomes more difficult when rare plants are involved. Recently, remarkable progress has been made toward identifying various categories of plant images based on deep learning [2–5]. Convolutional neural network (CNN) is a typical classification model [6]. Multilevel convolution and activation layers are used to extract high-dimensional features from raw images to simplify the final classification. Thus, CNN is more suitable for image processing than other traditional machine learning methods [7]. Several researchers have tried to explore the best structure or loss function, the misclassification error of which is not more than that of human experts [8–11]. However, training dataset is one of the major problems threatening classification accuracy. The construction of this model on an inadequate amount of data is susceptible to overfitting [12], and low-quality samples cause difficulty in learning effective features. Classic datasets (e.g., MNIST [13], CIFAR-10 [14], and ImageNet [15]) generally have thousands of image samples. However, the number of rare plant samples cannot achieve such scale, especially for rare samples. Traditional data enhancement methods, such as rotation or cropping, cannot easily produce new similar features [16]. They are more beneficial for resisting noise rather than improving classification. Gao et al. [4]

proposed an approach that developed a splicing method to generate synthetic sugar beet images. They first calculated the excess green (ExG) vegetation index [17] and normalized R, G and B channel. Afterwards, the ExG grayscale images were converted into binary mask images with Otsu's algorithm [18]. Finally, the object images and their masks were transformed using a set of randomly chosen parameters to synthesize final images. Unfortunately, only limited change has been done on the basis of original dataset.

A single-source small-sample image classification model based on latent semantics was proposed by Liu to address these problems [19]. Three semantic relationships of scene categories, images, and objects were analyzed initially. Then, potential semantic relationship was represented by calculated similarity, and the scenes in other categories, such as the less-category data, were searched. However, this method has relatively high demand for existing samples and requires additional annotation information, which are difficult to achieve in real-life scenarios. In addition, the calculation of similarity between images is unreliable because no perfect methods are currently available for measuring the distance between images. Although the Kullback–Leibler (KL) divergence can measure the distance between two distributions, the statistical error on a single image is extremely large. The structural similarity algorithm or mean square error is only available for symmetric data set.

Transfer learning is another broadly applicable method to solve small sample size problems [20–22]. The weight parameters are constrained to a good interval in the pretrained network. Thus, the model is expected to have low generalization error. If the original dataset used for training is sufficiently large and versatile, then the spatial hierarchy of the features of the pretrained network learning effectively serves as a general model of the visual world. Thus, the oversized dataset ImageNet [15] is generally used as the training dataset. Nevertheless, the probability that it will not work effectively on certain specific scenarios, such as medical images, is high. Identifying other datasets with similar characteristics is difficult because of the specificity of these sample data.

Generative adversarial network [23], as an emerging generative algorithm, has made

great progress in the quality and variety of generated images [24–36]. GAN can imitate the distribution of original data and output new samples with similar characteristics. It can fully use original datasets to dig out profound information. GAN is a well-recognized new data augmented method [37]. Valerio et al. [38] proposed a deep convolutional network that was able to generate synthetic rosette-shaped plants to dataset augmentation for plant phenotyping, inspired by DCGAN. The result showed that the testing error was reduced when the new samples were used as part of the training data on the state-of-the-art leaf counting algorithm. In this paper, we propose a new plant data augmentation method on the basis of an appropriative GAN model to aid classification. First, multilayer residual blocks are considered the backbone for enhancing the model's ability of expression and reducing trainable parameters. Few layers or parameters require few data sets. Second, we choose Wasserstein's divergence as a measurement level between two distributions. Hence, the generative network has output diversity and avoids mode collapse, thereby expanding raw datasets effectively. To guarantee the similarity between generated samples and the original samples in high-dimensional features, the content loss from Visual Geometry Group 16 (VGG16) [10] is added to the final loss. Therefore, the whole function pushes the solution to the natural image manifold. Lastly, the expanded dataset is tested on pretrained ResNet, a mainstream classification network, to ease the over-fitting phenomenon on the test set. The experiment shows that our method can solve plant small-scale classification through data augmentation better than other methods. Furthermore, considerable results verify that our algorithm has good migratory ability on other small sample datasets.

Methods

Prior research proved that rich features were the key to solve small-sample classification. Thus, this work focuses on fully using original small datasets to dig substantial valuable information. We introduce a new network for plant sample augmentation. The detailed architecture and optimization process of the proposed framework, namely, residual Wasserstein GAN will be clarified in the following subsection.

Structure of Res-WGAN

Res-WGAN is proposed as an enhanced version of the standard WGAN-gradient penalty (GP) [28] to leverage the success of GAN for image generation and application to small-scale plant datasets. It describes a new architecture that integrates 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 the generator is strengthened. Meanwhile, these connections introduce neither an extra parameter nor a computational burden. Content loss that guarantees the similarity between generated samples and original samples in high-dimensional features is effective for classification. More details about Res-WGAN are shown in Figure 1.

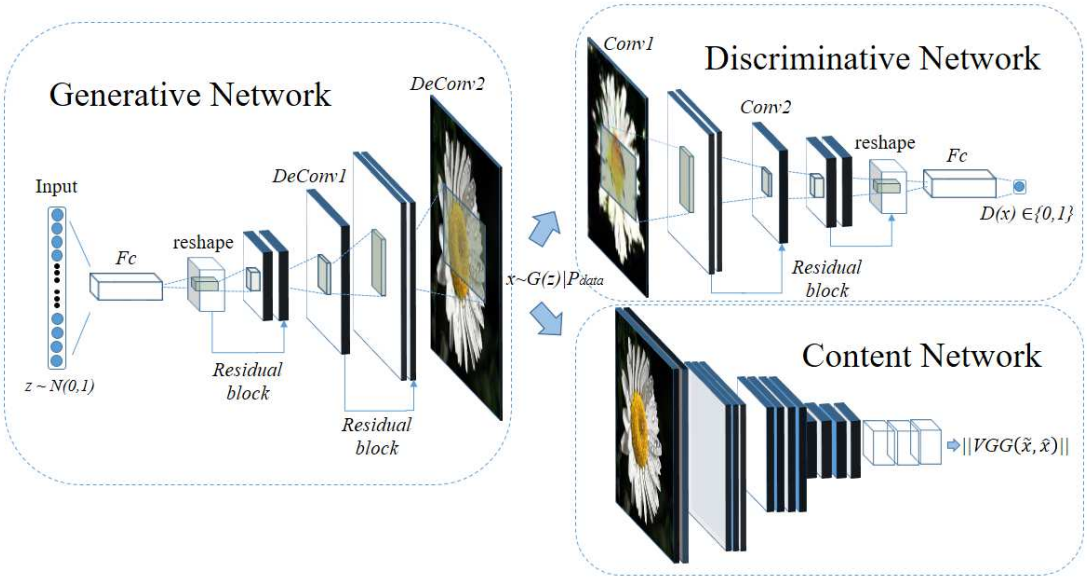


Fig. 1 Overview of our proposed Res-WGAN

It is predominantly composed of generative network, discriminative network, and content discriminator. The generative network, which is responsible for feature learning from random noise z to new sample $G(z)$, contains one fully-connected layer, two residual blocks, and two deconvolution layers. The discriminative network is made up of two convolution layers, one fully-connected layer, and several residual connections. Content discriminator is a substructure of the pretrained VGG model. These convolution layers are two parts of the whole discriminator. They 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, namely, the generative network (Figure 1), discriminative network, and content discriminator. The former is responsible for converting random noise z into images $G(z)$ similar to

sample x from the original dataset p_r visually. More specifically, z is a 100-dimensional random noise following the normal distribution, and it is projected to a small spatial extent convolutional representation with many feature maps. Then, a series of fractionally-strided deconvolutions converts this high-level representation into a 64×64-pixel image. The whole generator mainly contains two deconvolution layers (i.e., *DeConv1* and *DeConv2*), two residual blocks (i.e., *Res-DeConv1-1, Res-Conv1-2, Res-DeConv2-1,* and *Res-Conv2-2*), and one fully-connected layer (i.e., *Fc*). We use the basic architecture of ResNet and introduce 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 the basic output of the *DeConv2* layer and the result of the *Res-Conv2-2* layer. As shown in Figure 2, each block ends with a convolution layer whose stride is set to 1×1 to reduce the whole parameter count. More information about parameter settings are revealed in Table 1.

Table 1 Details of Res-WGAN layers

Layer	Type	Input size	Patch size	Kernel sum	Stride	Parameter
Generative Network						
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 detailed operation of the generative model complies with the original GAN

principle from Goodfellow [23]. The whole structure remains the same, but each part is enhanced to adapt to the plant small-scale datasets. The loss function of the generative network can be expressed as follows:

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

where L_G represents the whole loss of the generative network that consists of two parts, namely, 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 is sent into the generator G to obtain pictorial result $G(x)$. Then, the newly generated images are considered inputs of discriminator D to distinguish between real and fake, i.e., $D\left(G(x)\right)$. $\mathbb{E}$ is an expectation, which indicates that 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 pretrained VGG16 [10] 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. In the experiment, we randomly obtain eight images from the dataset p_r into the VGG [10] model and achieve the average output of the 13 layers as a content feature per iteration. To satisfy the input size, we enlarge every image from 64×64 to 224×224 artificially. μ represents the weight coefficient and is set to 0.001 normally.

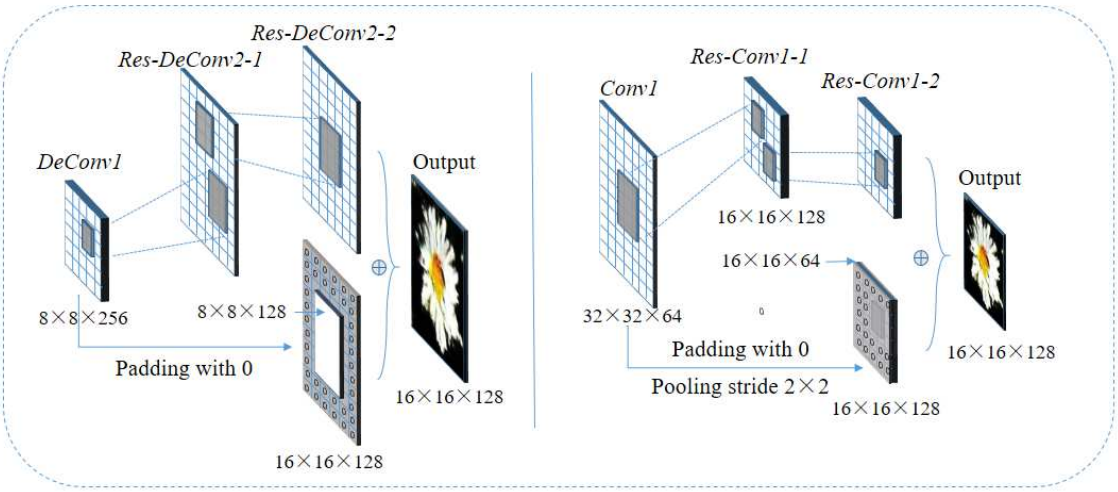


Fig. 2 Structure of residual blocks

The left part in Figure 2 is the second residual block in the 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 the *Res-Conv2-2* layer combined with the output of the *Deconv1* layer. To satisfy the size requirements, we consider zero as padding. Similarly, the first residual block in the discriminative network is shown on the right. Three convolution layers are available (i.e., *Conv1, Res-Conv1-1,* and *Res-Conv1-2*).

Figure 1 shows that another part of Res-GAN is the discriminator. The whole discriminator consists of a discriminative network and a content discriminator. A discriminative network is responsible for distinguishing whether the input is obtained from the ground truth or the generative network, that is, a two-class problem. It mainly contains two convolution layers (i.e., *Conv1* and *Conv2*), two residual blocks (i.e., *Res-Conv1-1, Res-Conv1-2, Res-Conv2-1,* and *Res-Conv2-2*), one fully-connected layer (i.e., *Fc*). More parameter information can be obtained from Table 1. Similarly, the residual block is introduced to enhance the ability of the discriminative network. For instance, the final output of the first residual block is equal to the sum of the basic output of the *Conv1* layer and the result of the *Res-Conv1-2* layer. A more specific connection is shown in Figure 2. A content discriminator is designed to calculate high-dimensional feature difference between two images, as described in SRGAN [36] with detail. Specifically, we obtain VGG16 [10], which is trained effectively on dataset ImageNet, as the content discriminator. The classification model is equipped with good generalization ability due to the abundance of species in ImageNet. The final loss function of the discriminative network is shown in Equation 3.

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

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. $D(x)$ indicates 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 [28]. This setting aims to establish a connection between the training gradient of the discriminative network and the hyperparameter K (usually set to 1). Thus, the function represented by the network satisfies the condition, that is, Lipschitz continuity. λ represents the weight coefficient, which is set to 10.0 normally. In conclusion, the final function of Res-WGAN can be expressed as follows:

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

Here, the ultimate objective is to minimize the generative loss while maximizing the discriminative loss. First, we initially consider the $\max D$ part while maintaining unchanged $\min G$. To the best of our knowledge, the goal of discriminative network D is to distinguish the input from the original dataset p_r or new samples. If sample x from the original dataset p_r is sent to D, then $D(x)$ is expected to approach 1. Second, x from random noise P_z is converted into image $G(x)$ by the generative network G; thus, $D\left(G(x)\right)$ is expected to approach zero. Third, as the gradient penalty $||\nabla_x D(x)||_p$ becomes closer to 1, the whole penalty term $||\nabla_x D(x)||_p - 1$ approaches 0. In summary, the overall loss, i.e., $\max D$ is expected to increase. In $\max D$ analysis, $\max D$ is shielded when $\min G$ is renewed. Only the second and fourth items are useful for the whole function. G is trained to spoof D to obtain $D\left(G(x)\right)$ that is close to 1. The fourth item represents the formula of content loss. It identifies the overall difference between original samples and new samples. The details of the training are shown in Table 2.

Table 2 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: Gradient penalty coefficient λ, number of critic iterations per generator iteration n_{critic}, batch size m, and Adam hyperparameters, namely, α, β_1, β_2.

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

while θ has not converged **do**

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

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

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

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

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

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

    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 $j = 1, \ldots, m$ **do**

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

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

  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

The tendency of classification accuracy to decrease as dataset decreases is a proverbial phenomenon. However, the baseline, where overfitting occurs, remains to be confirmed. The number of training samples was reduced gradually to visualize the accuracy change of the classification model. We chose ResNet-18 [11], which has less layers and is small sample friendly, as the classifier. To ensure the generality of results, a number of experiments have been conducted on different benchmark datasets. The results of "Flowers and Beans" are demonstrated in Figure 2 as examples. "Flowers" is a classic plant dataset, which contains five varieties of flowers, such as dandelion, sunflower, and tulip. Each species has approximately 1000 pictures. "Beans" contains 8000 images of soybean leaves under 10 different stresses. Each class in the test set has 100 samples

regardless of the size of the training set.

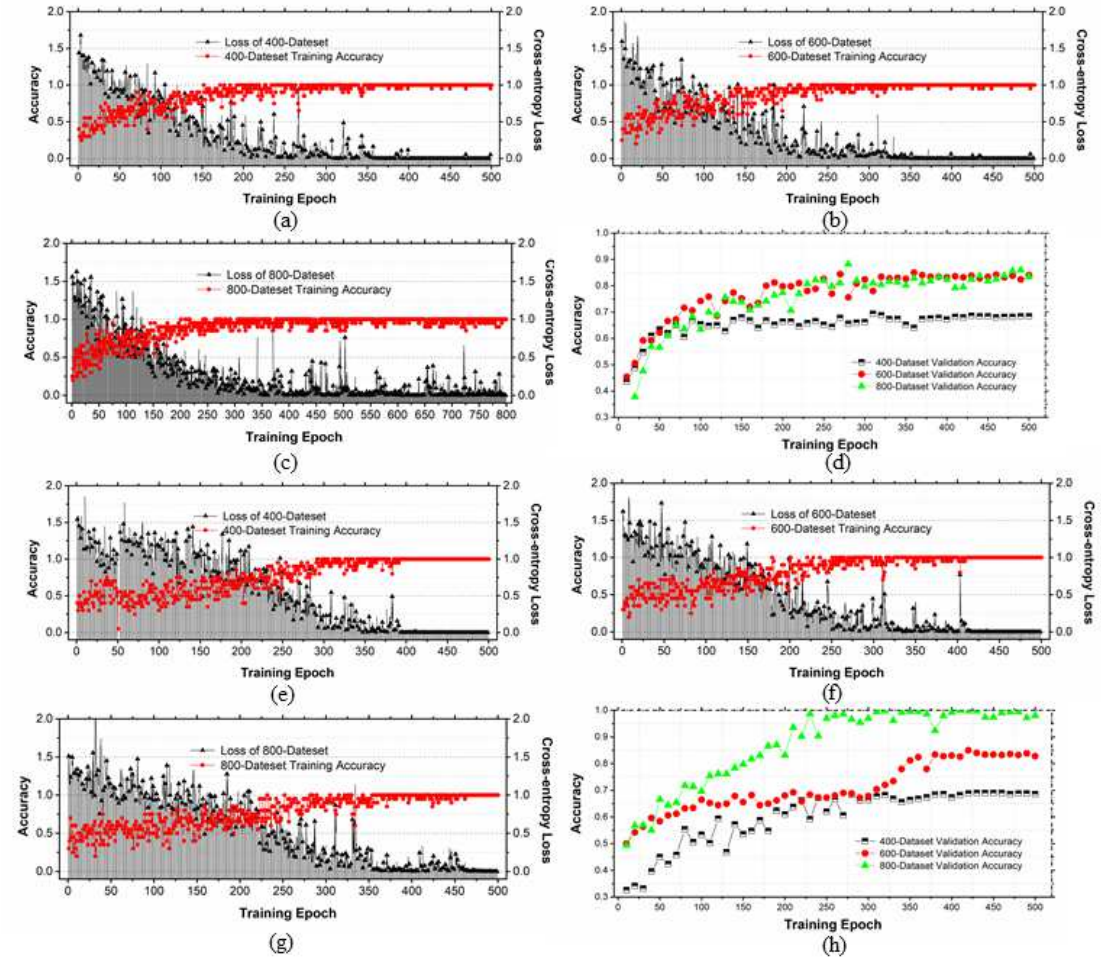


Fig. 3 Trend of training and validation accuracies varying with different dataset sizes. On the "Flowers" dataset, (a), (b), and (c) show the training accuracy when the number of single-type dataset is 400, 600, and 800, respectively; their validation accuracy is shown in (d). Similarly, (e), (f), (g), and (h) are the results of "Beans" dataset.

Figure 3 shows no significant decline in accuracy on training dataset with reducing numbers of training samples. The accuracy reaches nearly one and the loss reaches zero. The classification model ResNet [11] has learned good parameters by training. However, the accuracy of test data achieves great attenuation due to the decrease in training samples. This type of over-fitting phenomenon worsens when the single type of data set is 400. Many auxiliary strategies, such as Dropout and LReLu, have been used in training to prevent the problem. However, the classification model cannot effectively learn image features from actual small datasets.

Evaluations

Res-WGAN is adequately compared with other models or methods in terms of generative ability and accuracy improvement. The charts show that prediction on test data declines heavily when the sample size is set to 400 for each class. Thus, a new study is carried out on this basis.

Structure comparison

Time and space complexities are important indexes for evaluating algorithms. To prove that our method has not caused space consumption and that it is easy to train, we compare it with other state-of-the-art GANs, as shown in Table 3.

Table 3 Parameters, FLOPs, training epoch, and training time of models during training phase

Model	Parameter	FLOPs	Training Epoch	Training Time (h)	IS
Res-WGAN	25255621	304907093851	80	5.0	30.6
DCGAN	35701637	998396935281	70	4.0	18.5
WGAN	28328837	249636095476	70	3.5	21.2
WGAN-GP	28328837	395326772642	80	4.5	25.6
SAGAN	59257754	714406331436	80	5.5	31.5

Table 3 quantifies the comparisons of the main evaluations between Res-WGAN and state-of-the-art methods. Although the parameters decline into the lowest level in Res-WGAN compared with deep convolutional GAN (DCGAN) [24], WGAN-GP [28], and self-attention GAN (SAGAN) [30], more calculation amount and time are required for training in terms of floating point operations per second (FLOPs) and training time. Inception score (IS) is a common measurement index of quality and diversity of generated image. A total of 400 generated images of each class have been used to calculate IS. Our model ranks second at 30.6, following SAGAN [30] at 31.5. Moreover, SAGAN [30] is famous for generating high-quality images. All the models obtain lower score than training on large datasets because of the restriction of plant small-sample dataset.

Generative ability

All experiments have been completed on dataset “Flowers.” Visual images are strong proof of generative models. In addition to image sharpness and distortion, the diversity

of the output is another indispensable index to evaluate GAN. The outputs of DCGAN [24], WGAN-GP [28], SAGAN [30], and Res-WGAN with sufficient iteration are shown in Table 4.

Table 4 Outputs of different generative model training on flowers

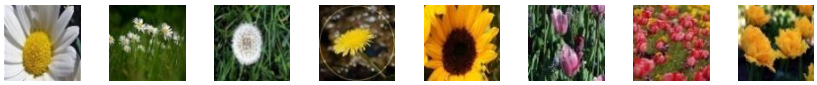
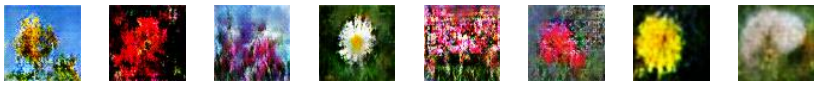
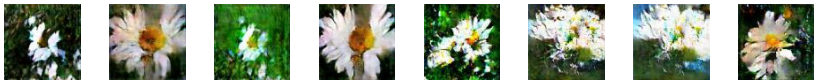
Model	Result
Raw Date	
DCGAN	
WGAN-GP	
SAGAN	
Res-WGAN	

Table 4 shows that the samples generated by the original WGAN-GP [28] and DCGAN [24] are blurry, and the texture details are disordered and missing. The expressive power of these complex networks is limited due to small sample training datasets. Thus, the network cannot effectively learn even with more training epochs. This finding can be confirmed by increasing the size of the training data. Our model produces better images due to its rich structure, as shown in Table 4. The images from Res-WGAN are clearer than the other results, and the details of texture are orderly as well. Moreover, these images are more easily classified visually. In addition, pattern collapse is observed in SAGAN [30] and DCGAN [24] in terms of the types of images generated. A more detailed description can be found in the study of raw WGAN-GP [28]. Pattern collapse is observed when Jensen–Shannon (JS) score is considered a loss function; it is replaced by KL divergence in our model. Thus, the results from our model have strong diversity and are beneficial to the next classification training.

Classification improvement

The expanded dataset is used for the training set of ResNet [11] to obtain better

classification accuracy. To fully verify the effectiveness of enhancing classification capacity, other general enhancement methods, such as rotation [39], random erasing [40], cropping, and transfer learning, have been added into the experiment (Table 5).

Table 5 Accuracy of training and validation on datasets with different sizes and 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**

Table 5 shows that the training accuracy already approaches one when the sample size of each class is set to 400. Thus, the duplicate results are removed when the size of the dataset is 600 or 800. A step further is difficult because of the training bottleneck. Moreover, the dataset has poor-quality images that are difficult to classify. Improving the model's accuracy on test data is the key point in this study. Overfitting is often observed in small-sample classification tasks, as shown in Tables 2 (d) and (h), although some countermeasures, such as dropout or leaky ReLU, have been used. Traditional data enhancement improves over-fitting compared with other methods. When the size per category in the original dataset is expanded from 400 to 600 or 800, the accuracy on ResNet improves by only 1.3%. Although these new samples promote translational and rotational invariance of the model, they are still a subset of the original set.

Different from data enhancement methods, transfer learning updates parameters by pretraining on other datasets and uses a small sample set for fine-tuning. Compared with other single methods, transfer learning obtains the highest validation accuracy (82.9%). Corresponding to transfer learning, the purpose of the proposed method is to fundamentally learn high-dimensional features from various image types and generate similar features, which in turn help the classification model to learn. In the experiment,

a one-to-one training mode is considered to produce new samples. The same types of flowers are generated one by one. The size of these images is fixed at 64×64, which is smaller than the image size in the original dataset but consistent with the input of the discriminative network. Large images are difficult to output because of the restriction of the small-scale samples. Finally, the combined dataset (generated images and original images) is used to train ResNet [11], and the best validation accuracy can reach 75.1%, which is the third highest accuracy in augmented datasets (Table 5). The result confirms that our proposed Res-WGAN model can effectively learn features and enhance the generalization ability of the classification model.

Lastly, inspired by transfer learning, we use a pretrained model as a base model and continue to train the extended dataset. The validation accuracy (83.7%) exceeds all results for the same period. However, this result is consistent only with actual training dataset size equal to 600. A gap from the best results is still found. Compared with DCGAN and SAGAN, the WGAN [27] series performs better on small-scale plant datasets. This finding is probably related to the model's generative capacity. The relationships between them will be studied in depth in our future work.

Model migration

Various forms of scenes are observed in botanical research, where our model can play an important role. The Res-WGAN model is used for other small datasets to verify its generalization ability. The dataset includes three types of green plants, including *Arabidopsis*, beans, and Romaine lettuce. Each class contains 400 image samples. Accurate classification in the actual experimental environment is necessary. In accordance with the result of a previous study, we expand the original dataset from 400 images in each class to 800 by using Res-WGAN. Then, they were inputted to ResNet [11] to test the accuracy. Table 6 shows the enhanced dataset obtained by Res-WGAN.

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

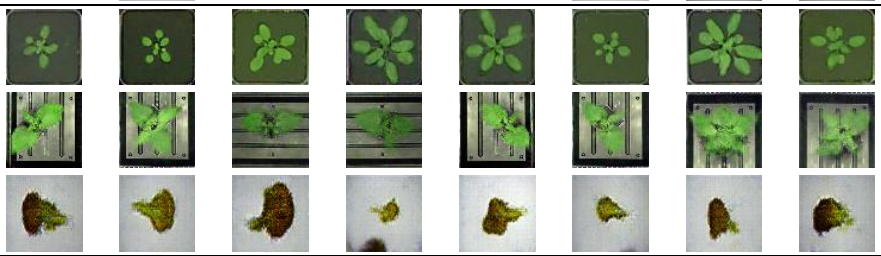
Raw data	
Result	

Table 7 Training and validation accuracy on datasets with different sizes and 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 obtains the best result among the other state-of-the-art methods (Table 7). This finding is consistent with the results in Section 4.2.3. The result is only 1.8% less than the actual dataset with 800 images in each class. It confirms that new samples generated by Res-WGAN enhance the generalization ability of the classification model. The proposed method is a favorable alternative to the original data set when the actual size of the dataset is limited.

Discussion

The original dataset is a constraint in the automated classification of rare plant samples using CNN. At the beginning of the experiments, we explore the impact of training samples at different levels on the classification model. Two datasets are selected, namely, “Flowers” and “Beans,” to reflect objectivity. Furthermore, we have selected three orders of magnitude of 400, 600, and 800 with significant changes in accuracy. They are used to identify the current number of samples of each class that participates in the classification. The classification model has already experienced a serious

overfitting phenomenon when the number is set to 400. If we continue to reduce the number, then this phenomenon will become more serious and the research will have low significance. Therefore, in the next experiment, we set the basic data size to 400, which is convenient for comparing the effects of different enhancement methods on classification. Unfortunately, the number of rare plant images sometimes fail to reach a magnitude of 400. However, as an exploratory method, our work solves the classification problem within this scale effectively. Future research should study the classification of smaller plant samples.

We have proven the superiority of the proposed method in multiple angles. The model expressibility is enhanced by reducing the network layer and introducing residual connections. Meanwhile, the number of parameters is reduced. Therefore, the training of the model has reduced its dependence on data. Thus, it is effective for rare plant data. However, the training time increases because the residual structure introduces more calculations. In terms of the quality of newly generated samples, Table 4 shows a good comparison. From a visual aspect, the generated data obtained by Res-WGAN have clear texture features and can effectively avoid model collapse. This finding is mainly attributed to the loss function, which has been described in detail in the previous section. The improvement in classification tasks is investigated by comparing the proposed method with different enhancement methods. Transfer learning is the most efficient method for a single method. The accuracy of the testing data can be improved from 0.682 to 0.829, which is close to the original sample size of 600 (0.835), by using pretrained parameters on other datasets. Our method ranks second (0.751). The results confirm that the Res-WGAN improves the accuracy of the classification model. At the same time, the content loss guarantees the similarity between generated samples and original samples in high-dimensional features. This concept is effective for final classification. The table shows that Res-WGAN can reach the accuracy of 0.746, which is 0.04 more than the best WGAN-GP [28] in the classic GAN and 0.06 more than the traditional data enhancement method, when the size of the training data is expanded to 600.

Nevertheless, the result when using only GAN for data enhancement is unsatisfactory.

Thus, we combine two state-of-the-art methods and obtain the result of 0.832, which is almost equal to the result when the original sample is 600. However, when the amount of training data is expanded to 800, the accuracy is not improved significantly. This finding indicates that this type of enhancement method has reached the bottleneck. More new image samples are ineffective for classification and introduce noise, as reflected in DCGAN and WGAN-GP. Finally, to prove the robustness of our model on other datasets. We verified our model on more datasets. The new experimental results fully satisfy our expectations.

Res-WGAN, which is effective in image generation, still has some shortcomings. We failed to successfully generate large-scale sample data using the proposed model. The newly generated samples did not form complete texture features but considerable noise. We assume that the small sample training data restrict the expressibility of the model. Nevertheless, large-scale image samples cover more pixels, and studying the relationship between pixels becomes more difficult. This finding proves that certain constraints exist on the application of our model. If the amount of original sample data is extremely small, then the model's generating ability decreases significantly. Improving the quality of images from GAN by using small-scale samples is important in our future work.

Conclusion

Special techniques are required to expand available datasets for effective classification due to the inaccessibility of rare plant samples in actual practice. An effective data augmentation method based on a dedicated GAN is proposed. Species of rare plants are accurately identified using transfer learning. The main contributions and shortcomings are summarized as follows.

A novel generative model, Res-WGAN, which is suitable for plant study especially for small sample size, is proposed. The new structure selects residual block as an elementary building unit and adds perceptual loss into Wasserstein's loss for rare plant samples. The experimental results showed that the generated samples of Res-WGAN have better performance in image quality and diversity than other methods or models. The overfitting phenomenon of ResNet [11] on small sample size data remarkably

improved by using transfer learning method. This finding provides a possibility of using a limited amount of data to train the classification model with high accuracy. Our model should also be applied to other scenarios with small sample size (e.g., medical images). However, our work still has some shortcomings. Generated sample quality is also limited by raw datasets, although the size of these images is set to be as small as possible. Furthermore, the sample size of our study begins at 400; a smaller sample size will cause the generator to fail. Our next research work will focus on finding methods to overcome the texture distortion in new sample images. An expanded dataset will be developed for effective classification. We aim to investigate the potential value of available samples and create an automatic classification tool to replace the high-threshold plant identification task.

Supplementary information

Abbreviations

CNN: convolutional neural network

GAN: generative adversarial network

Res-WGAN: residual wasserstein GAN

KL: Kullback–Leibler

GP: gradient penalty

VGG16: visual geometry group 16

ExG: excess green

SRGAN: super-resolution generative adversarial network

DCGAN: deep convolutional generative adversarial network

SAGAN: self-attention generative adversarial network

FLOPs: floating point operations per second

IS: inception score

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, DY, 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 final 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 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 and Rural Affairs, Beijing 100083 China

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

*Corresponding author: Minjuan Wang and Lihua Zheng

References

[1] Stuessy, Tod F. Plant taxonomy: the systematic evaluation of comparative data. Columbia University Press, 2009.

[2] Lin, P., Li, D., Zou, Z., Chen, Y., & Jiang, S. Deep convolutional neural network for automatic discrimination between Fragaria× Ananassa flowers and other similar white wild flowers in fields. Plant

methods, 2018; 14(1), 64.
[3] Ravindran, P., Costa, A., Soares, R., & Wiedenhoeft, A. C. Classification of CITES-listed and other neotropical Meliaceae wood images using convolutional neural networks. Plant methods, 2018; 14(1), 25.
[4] Gao, J., French, A. P., Pound, M. P., He, Y., Pridmore, T. P., & Pieters, J. G. Deep convolutional neural networks for image-based Convolvulus sepium detection in sugar beet fields. Plant Methods, 2020; 16(1), 1-12.
[5] Šulc, M., & Matas, J. Fine-grained recognition of plants from images. Plant Methods, 2017; 13(1), 115.
[6] 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.
[7] Kim, J. Y., Lee, H. E., Choi, Y. H., Lee, S. J., & Jeon, J. S. CNN-based diagnosis models for canine ulcerative keratitis. Scientific reports, 2019; 9(1), 1-7.
[8] Krizhevsky, A., Sutskever, I., Hinton, G. ImageNet Classification with Deep Convolutional Neural Networks. In Advances in neural information processing systems, 2012; 1097-1105.
[9] 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.
[10] Simonyan, K., & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv. 2014; 1409-1556.
[11] 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.
[12] Li, F., Chen, H., Liu, Z., Zhang, X. D., Jiang, M. S., Wu, Z. Z., & Zhou, K. Q. Deep learning-based automated detection of retinal diseases using optical coherence tomography images. Biomedical Optics Express, 2019; 10(12), 6204-6226.
[13] Deng, L. The mnist database of handwritten digit images for machine learning research [best of the web]. IEEE Signal Processing Magazine. 2012; 29(6):141-142.
[14] Krizhevsky, A., & Hinton, G. Learning multiple layers of features from tiny images. Technical report, University of Toronto. 2009. p. 7.
[15] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition. 2009. p. 248-255.
[16] 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.
[17] Otsu N. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans Syst Man Cybern. 1979; 9:62–6.
[18] Farid H. Blind inverse gamma correction. IEEE Trans Image Process. 2001; 10:1428–33.
[19] Liu, C.W. Small Sample Image Classification Based on Semantic Computation. Doctor Thesis of Chongqing University, Chongqing, China, 2017.
[20] Song, G.H. Image Annotation Method based on Transfer Learning and Deep Convolutional Feature. Doctor Thesis of Zhejiang University, Hangzhou, China, 2017.
[21] Wu G.Q. Research and Application of Transfer Learning in Image Classification. Master Thesis of Anhui University, Anhui, China, 2017.
[22] 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.

[23] 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.
[24] Radford, A., Metz, L., & Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. arXiv preprint, arXiv:1511.06434. 2015.
[25] 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.
[26] 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.
[27] Arjovsky, M., Chintala, S., Bottou, L. Wasserstein GAN. arXiv preprint, arXiv: 1701.07875. 2017.
[28] Gulrajani, I., Ahmed, F., Arjovsky, M., et al. Improved training of wasserstein gans. Advances in neural information processing systems, 2017;5767-5777.
[29] Miyato, T., Kataoka, T., Koyama, M., & Yoshida, Y. Spectral normalization for generative adversarial networks. arXiv preprint. arXiv:1802.05957. 2018.
[30] Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. Self-attention generative adversarial networks. arXiv preprint. arXiv:1805.08318. 2018.
[31] Brock, A., Donahue, J., & Simonyan, K. Large scale gan training for high fidelity natural image synthesis. arXiv preprint. arXiv:1809.11096. 2018.
[32] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision. 2017;2223-2232.
[33] 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.
[34] Miyato, T., & Koyama, M. cGANs with projection discriminator. arXiv preprint. arXiv:1802.05637. 2018.
[35] 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.
[36] 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.
[37] Yang, Y. N., Qi, L. H., Wang, H., & Su, L. P. Research on Generation Technology of Small Sample Data Based on Generative Adversarial Network. Electric power construction. 2019;71-77.
[38] 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.
[39] 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.
[40] Zhong, Z., Zheng, L., Kang, G., Li, S., & Yang, Y. Random erasing data augmentation. arXiv preprint.arXiv:1708.04896. 2017.

Figures

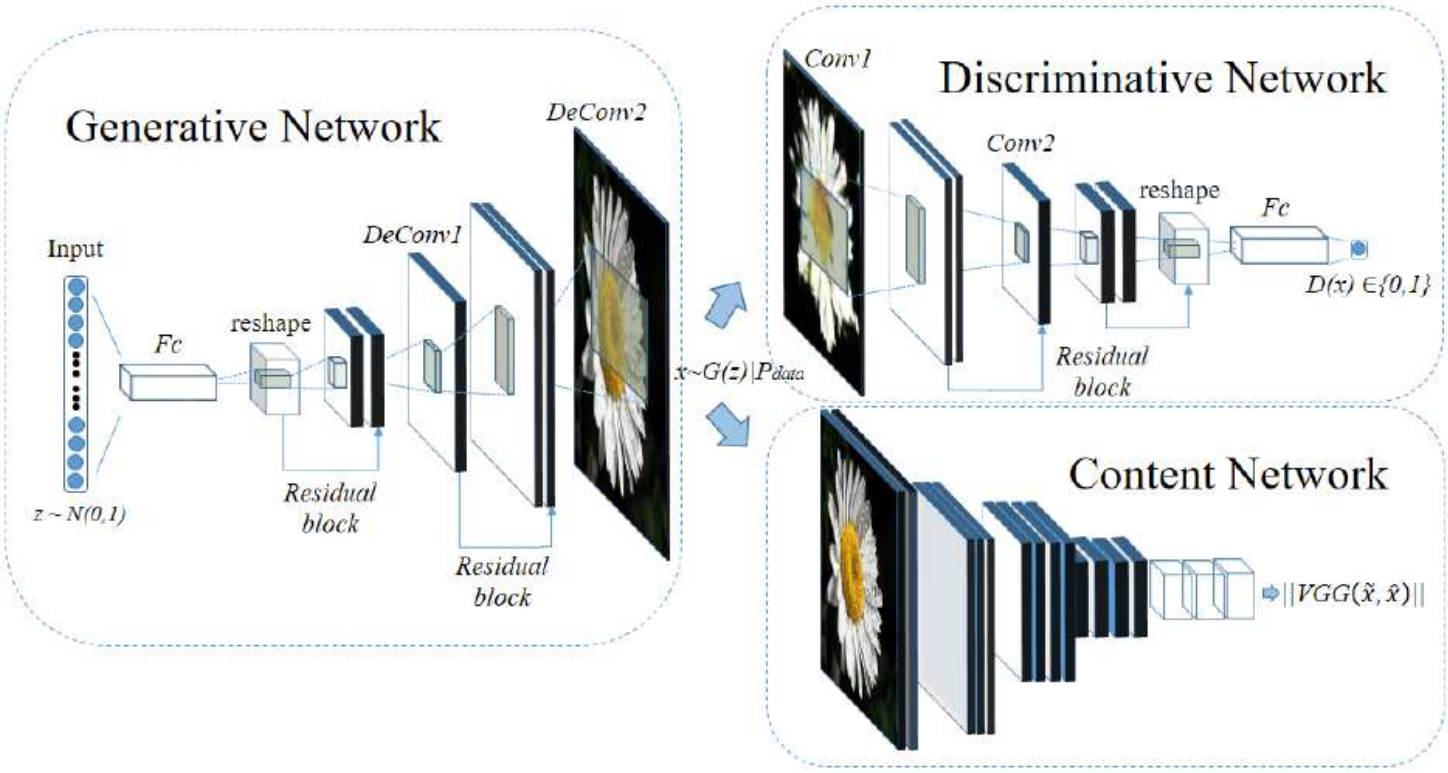


Figure 1

Overview of our proposed Res-WGAN. It is predominantly composed of generative network, discriminative network, and content discriminator. The generative network, which is responsible for feature learning from random noise □□ to new sample □□(□□), contains one fully-connected layer, two residual blocks, and two deconvolution layers. The discriminative network is made up of two convolution layers, one fully-connected layer, and several residual connections. Content discriminator is a substructure of the pretrained VGG model. These convolution layers are two parts of the whole discriminator. They distinguish whether the inputted image is derived from the original dataset □□□□ or generated images □□(□□).

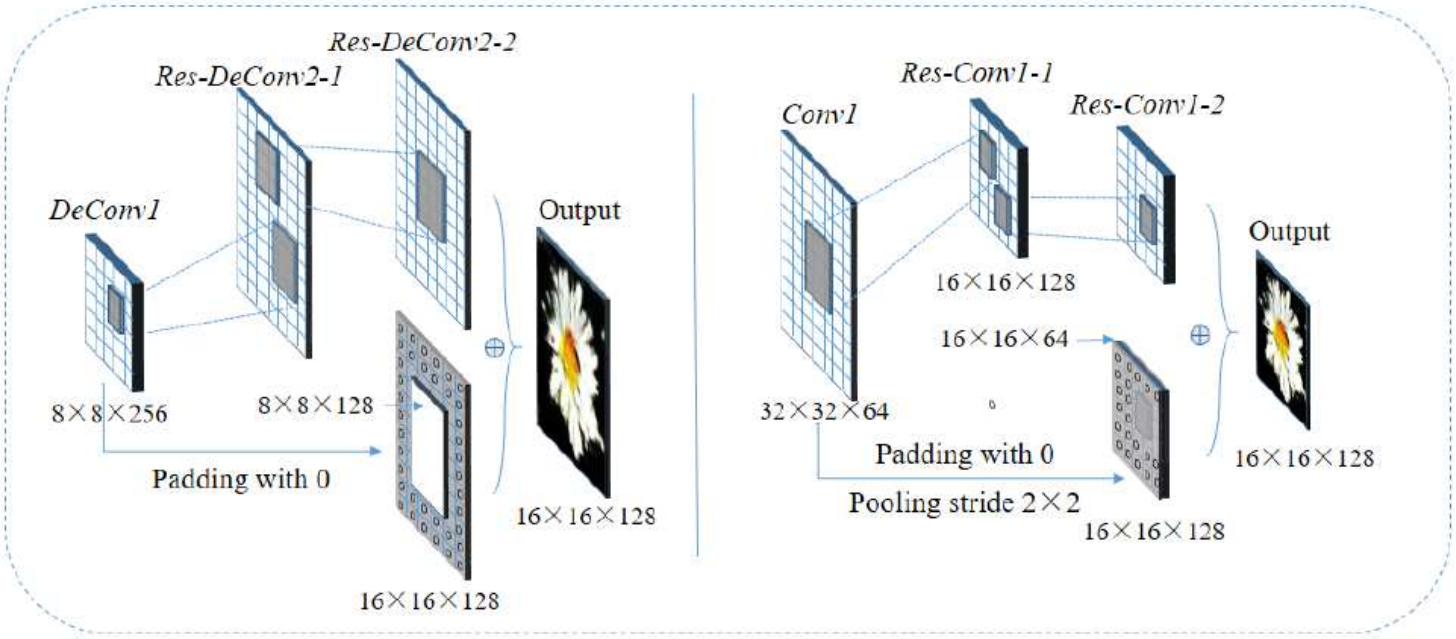


Figure 2

Structure of residual blocks

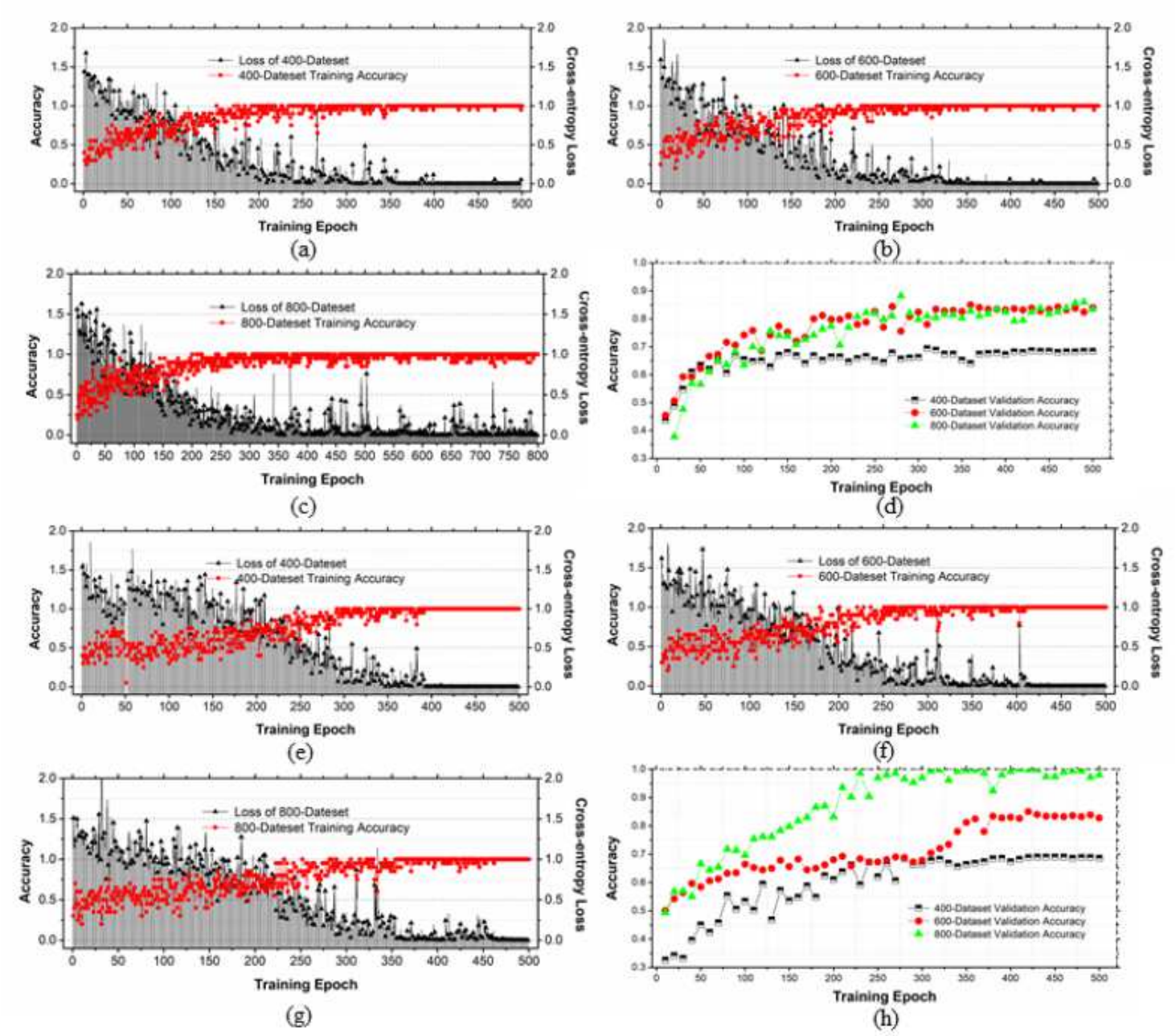


Figure 3

Trend of training and validation accuracies varying with different dataset sizes. On the “Flowers” dataset, (a), (b), and (c) show the training accuracy when the number of single-type dataset is 400, 600, and 800, respectively; their validation accuracy is shown in (d). Similarly, (e), (f), (g), and (h) are the results of “Beans” dataset.