Machine Vision and Applications (2024) 35:22
https://doi.org/10.1007/s00138-023-01506-y

ORIGINAL PAPER

SGBGAN: minority class image generation for class-imbalanced datasets

Qian Wan[1] · Wenhui Guo[1] · Yanjiang Wang[1]

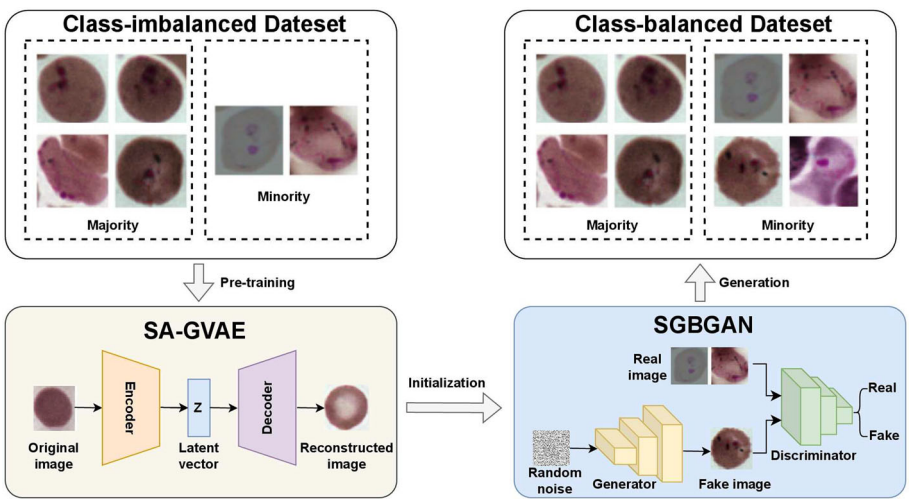
Received: 31 March 2023 / Revised: 25 August 2023 / Accepted: 20 December 2023 / Published online: 29 January 2024

Abstract

Class imbalance frequently arises in the context of image classification. Conventional generative adversarial networks (GANs) have a tendency to produce samples from the majority class when trained on class-imbalanced datasets. To address this issue, the Balancing GAN with gradient penalty (BAGAN-GP) has been proposed, but the outcomes may still exhibit a bias toward the majority categories when the similarity between images from different categories is substantial. In this study, we introduce a novel approach called the Pre-trained Gated Variational Autoencoder with Self-attention for Balancing Generative Adversarial Network (SGBGAN) as an image augmentation technique for generating high-quality images. The proposed method utilizes a Gated Variational Autoencoder with Self-attention (SA-GVAE) to initialize the GAN and transfers pre-trained SA-GVAE weights to the GAN. Our experimental results on Fashion-MNIST, CIFAR-10, and a highly unbalanced medical image dataset demonstrate that the SGBGAN outperforms other state-of-the-art methods. Results on Fréchet inception distance (FID) and structural similarity measures (SSIM) show that our model overcomes the instability problems that exist in other GANs. Especially on the Cells dataset, the FID of a minority class increases up to 23.09% compared to the latest BAGAN-GP, and the SSIM of a minority class increases up to 10.81%. It is proved that SGBGAN overcomes the class imbalance restriction and generates high-quality minority class images.

Graphical abstract

The diagram provides an overview of the technical approach employed in this research paper. To address the issue of class imbalance within the dataset, a novel technique called the Gated Variational Autoencoder with Self-attention (SA-GVAE) is proposed. This SA-GVAE is utilized to initialize the Generative Adversarial Network (GAN), with the pre-trained weights from SA-GVAE being transferred to the GAN. Consequently, a Pre-trained Gated Variational Autoencoder with Self-attention for Balancing GAN (SGBGAN) is formed, serving as an image augmentation tool to generate high-quality images. Ultimately, the generation of minority samples is employed to restore class balance within the dataset.



Keywords Autoencoders · Class imbalance · Generative adversarial networks · Gated linear unit · Self-attention

1 Introduction

1.1 Background and motivations

Image classification and image segmentation have always been classic topics in computer vision [1]. Training a deep neural network usually requires large and balanced datasets, such as Fashion-MNIST [2], CIFAR-10 [3], ImageNet [4], and others, they can be used to train modern image classifiers. However, the class distribution of existing datasets is often unbalanced, especially in the medical field. Due to the rarity and difficulty of collecting cancerous images, there are usually more normal images. Training deep neural networks with unbalanced datasets will result in poor model performance, and the results tend to be biased toward most classes [5]. Manually annotating to build a large and balanced labeled dataset is time-consuming and virtually impossible. Therefore, an effective solution is to balance the categories of the dataset to generate high-quality new minority class images [6].

Generative adversarial networks (GANs), as one of the popular generative models, are an effective method to increase minority class samples in image generation [7]. The main structure of a GAN consists of a generator and a discriminator. Through the adversarial training, the discriminator can judge whether the output result of the generator is true or not, so as to generate high-quality samples. However, the image generation based on GAN also has some shortcomings. It often requires large and balanced datasets, and the results tend to be biased to the majority class samples. Moreover, there are also some problems, such as mode collapse and second-best initialization, which lead to unstable data results [8].

The gated linear unit (GLU) is a gating mechanism that decouples input features and applies affine transformation and nonlinear activation, enabling flexible combination of linear mapping and nonlinear activation [9]. By providing element-wise gating mechanisms, GLU can enhance the nonlinear mapping capability of GAN generators, improving model expressiveness and generating sample quality for GANs [10]. Self-attention mechanism aggregates value vectors with weights calculated from relevance between queries and keys, to obtain representations focused on different positions in the input. By modeling global dependencies, self-attention enhances capability to model long-range context, enabling GAN to generate more coherent and consistent samples [11].

Feature extraction is an important step in image classification tasks. Dimensionality reduction through feature extraction can reduce data redundancy, reduce computational complexity, and make the classification process more efficient. A new feature extraction method based on rotation invariant neighborhood binary patterns is applied to block-based copy-move forgery detection, which improved the detection performance [12]. In order to train GANs on class-imbalanced datasets, recent works combine GANs with autoencoders. The powerful feature extraction and reconstruction capabilities of autoencoders are used to enhance the initialization and training of GAN-based models to generate minority class samples. Balancing GAN (BAGAN) [13] and BAGAN with gradient penalty (BAGAN-GP) [14] are tools to deal with datasets with class imbalance, and good results are obtained when solving some problems caused by class imbalance. Although the latter are an improvement on the former, they are still limited by the problem of category imbalance. First, BAGAN-GP uses an ordinary autoencoder and an objective function for initialization training. Better initialization could be obtained by changing the initialized autoencoder architecture to generate minority class images with higher quality. In addition, in the process of adversarial training, GAN also uses common architecture, which cannot adapt to complex data distribution. Moreover, they only conduct comparative experiments of different methods on a single dataset and lack assessment of different imbalance levels in other datasets.

1.2 Innovations

The high degree of freedom of the potential space is a problem because it can lead to serious over-fitting without information loss, and some vectors of the potential space become meaningless information when decoded. For this reason, ensuring the regular performance of the potential space is enough to enhance the decoder generation capability of the autoencoder. The possible way to learn this regularity is to introduce regularization in the training process. The variational autoencoder introduces explicit regularization in the training process to avoid over-fitting in the decoding process and improve the decoder generation ability [15]. Therefore, we use a variational autoencoder with regularization for pre-training, effectively reducing the over-fitting problem of the model. Adding a self-attention mechanism to the encoder and decoder to weight image details texture features, so as to better extract features. In addition, GLU is added to the convolutional layer structure of the encoder to accelerate the convergence rate of the network and slow down the dis-

✉ Yanjiang Wang
yjwang@upc.edu.cn

Qian Wan
1322520157@qq.com

Wenhui Guo
whguo3823269@163.com

[1] College of Control Science and Engineering, China University of Petroleum (East China), Qingdao 266580, Shandong Province, People's Republic of China

appearance of gradient. A Gated Variational Autoencoder with Self-attention (SA-GVAE) is proposed to perform initial pre-training on the GAN. The weight of the pre-trained autoencoder is transferred to GAN to obtain a good starting point for adversarial training, namely Pre-trained Gated Variational Autoencoder with Self-attention for Balancing GAN (SGBGAN), which could generate high-quality minority class samples for small and class-imbalanced datasets, address the limitations of imbalance.

1.3 Contributions

Our contributions can be summarized as follows:

- We introduce a self-attention mechanism and GLU into the variational autoencoder, and propose a SA-GVAE which has strong feature extraction and reconstruction capabilities to learn the distribution of real samples;
- We combine the proposed SA-GVAE model with GAN, then initialize and train the corresponding GAN components to enhance the training stability, namely SGBGAN;
- We propose a new complex objective function to encourage the model to capture the true distribution of samples and produce high-quality samples, the model could cope with complex sample distribution;
- We use different imbalance rates on three different datasets for a more comprehensive assessment to generate more realistic and diversified samples.

2 Related work

Generative models deal with class imbalance. At present, the most classical generative models are autoencoders and GANs. Using them to solve the class imbalance restriction has been a hot topic in recent years. Studies have shown that in addition to reducing the majority of class samples, generating more minority class samples is the most useful method to restore the class balance of data. A few studies use autoencoders and their variants to solve the class imbalance problem. For example, Taghanaki et al. propose a variational autoencoder regularization scheme to solve the feature imbalance problem [16]. Zheng et al. introduce the GLU to enhance the feature extraction capability of encoders [17]. However, Li et al. find that the diversity of the images generated by the variational autoencoder is poor, while the images generated by the GAN are improved [18]. Therefore, most researchers have discovered the powerful generation ability of GAN and its advantages in processing class-imbalanced datasets, they start to use GAN and its variants to generate minority class samples. Shoohi et al. propose that Conditional Deep Convolutional Generative Adversarial Network (DCGAN) can generate new images for minority class samples and produce high-quality results on a variety of image generation tasks [19]. Wang et al. prove that the Wasserstein GAN could learn the basic distribution of a few types of samples and use oversampling technology to enhance computed tomography image data [20]. Zhu et al. propose Cycle GAN, they apply image-to-image conversion to datasets with class imbalance, and try to generate minority class samples from most class samples [21]. In recent years, semi-supervised and unsupervised GANs have been used to achieve class-imbalanced image generation. Balasubramanian et al. use an unconditional GAN to perform image enhancement on cancerous data of diabetic retinopathy [22]. Moreover, Waheed et al. propose CovidGAN for the amplification of a few Covid-19 Chest X-Ray images that are difficult to collect [23]. Although model generation results are good when facing the problems caused by class imbalance, there are still some problems such as mode collapse and unstable training. In addition, Sampath et al. believe that the quality of minority samples generated by GAN is not good enough and will reduce the classification performance of the original classifier; the result is more favorable to the majority of classes [24].

Data augmentation based on GAN and autoencoder. Traditional image data augmentation methods such as flip, crop, scale, translation and so on are just simple geometric transformation. It is difficult to make the model learn more useful information from the original dataset [25]. In order to extract more feature information from the original dataset and overcome the restriction caused by class imbalance, the autoencoder is combined with GAN to carry out nonlinear image generation to obtain high-quality minority class samples, such as BAGAN. It initializes GAN through pre-trained autoencoder and provides a more stable starting point for GAN training on the basis of better feature extraction and reconstruction capabilities [13]. However, the use of unsupervised autoencoders makes it impossible to classify images by category when transforming them into latent vectors, so label information is very important for the pre-training stage of autoencoders [26]. Zheng et al. propose a conditional Wasserstein GAN integrated with gradient penalty regularization (cWGAN-GP) to generate synthetic minority class data for balancing class-imbalanced datasets [27]. Then, BAGAN-GP is developed, which is improved on BAGAN. It believes that BAGAN has poor results on small-scale medical datasets, and the results are biased toward most categories [14]. BAGAN-GP uses an autoencoder with an embedding model to learn label information in the pre-training stage, and also applies the conditional GAN to specify categories in the adversarial training process of GAN to generate high-quality minority class images. The results on manually created unbalanced versions of the dataset also demonstrate the superiority of BAGAN-GP over BAGAN. Dai et al. propose a framework for enhancing and classify-

Table 1 Related literature summary and comparison

Method	Innovation and contribution
DCGAN	Generates new images for minority class samples but the quality is poor
CycleGAN	Applies image-to-image conversion to datasets with class imbalance
CovidGAN	Generates hard-to-collect Covid-19 CXR images
BAGAN	Initializes GAN through pre-trained unsupervised autoencoder
cWGAN-GP	Integrates with gradient penalty regularization
BAGAN-GP	Uses an autoencoder with embedding model to learn label information
BWGAN-GP	Handles class imbalance in RSVP tasks and generates artificial EEG data

ing spot welding defect images on the basis of BAGAN-GP, and the resulting images are added to the training dataset to improve the classifier for better classification performance [28]. Xu et al. propose balanced Wasserstein generative adversarial network with gradient penalty (BWGAN-GP), which can handle class imbalance problems in RSVP tasks and generate artificial EEG data [29]. Ding et al. propose a forward and reverse design platform based on BAGAN-GP for the design of cross-ribosome binding site for transcription factor-based biosensors, which performs well in processing class-imbalanced datasets [30]. The literature review and comparison are shown in Table 1. However, BAGAN-GP uses ordinary autoencoders in the pre-training stage, its simple architecture and objective function, as well as the pre-training stage with imbalanced data, all make the results more favorable to majority classes, and the training results are not stable enough.

3 Methods

3.1 Improved Gated Variational Autoencoder with Selfattention in SGBGAN

Similar to the common autoencoder, the purpose of variational autoencoder training is to minimize the reconstruction error between the reconstructed data and the original data. However, instead of outputting a single vector of the potential space, the encoder learns the probability distribution of the potential space. Although variational autoencoder has achieved good results in many tasks, there are still some problems such as prior collapse and low image quality, especially the image with blurred edges [31].

A Gated Variational Autoencoder with Self-attention (SAGVAE) is proposed to further improve the quality of model generated images. The self-attention mechanism is introduced into the variational autoencoder model for the purpose of weighted image detail texture features, so as to better feature extraction, the network flowchart of the self-attention mechanism layer is shown in Fig. 1. Among them, $\boldsymbol{x} \in R^{c*w*h}$, $\bigotimes$ refers to matrix elements multiplication. After two 1×1 convolutions of the input convolution feature graph $\boldsymbol{x}$, the feature graph is obtained. The matrices of the two feature graphs are multiplied to calculate the attention diagram [11]. The calculation formula is shown in Eq. 1.

$$\beta_{j,i} = \frac{\exp\left(s_{ij}\right)}{\sum_{i=1}^{N} \exp\left(s_{ij}\right)} \tag{1}$$

where $\beta_{j,i}$ represents the degree of attention of the model to the i-th position when the j-th region is synthesized, $s_{ij} = \boldsymbol{f}\left(\boldsymbol{x_i}\right)^T \boldsymbol{g}\left(\boldsymbol{x_j}\right)$. After multiplying the attention map and the matrix of the feature graph, a 1×1 convolution is carried out to restore the number of channels to the same number as the number of input channels, and the self-attention feature map $\boldsymbol{o}$ is obtained. The calculation formula is shown in Eq. 2.Then, the scale parameter γ is used for scaling operation to obtain the output y. The calculation formula is shown in Eq. 3. Starting with local information and moving on to distant dependencies. For computer vision tasks, the advantage of adding attention mechanism is that the model can be more intuitively understood [32].

$$\boldsymbol{o}_j = \boldsymbol{v}\left(\sum_{i=1}^{N} \beta_{j,i} \boldsymbol{h}\left(\boldsymbol{x}_i\right)\right) \tag{2}$$

$$\boldsymbol{y}_i = \gamma \boldsymbol{o}_i + \boldsymbol{x_i} \tag{3}$$

In addition, the structure of the convolutional layer of the encoder in the variational autoencoder is adjusted, and the GLU is added to the convolutional layer. Its network is shown in Fig. 2. The results of the convolutional layer are divided according to half of the number of convolutional kernels, one part of which is taken as the data feature and the other as the weighted matrix, and then the data feature and the weighted matrix are multiplied element by element [33]. The output of the partial activation layer is obtained, and the calculation formula is shown in Eq. 4.

$$\boldsymbol{h}(\boldsymbol{x}) = (\boldsymbol{x} * \boldsymbol{W} + \boldsymbol{b}) \otimes \sigma(\boldsymbol{x} * \boldsymbol{W} + \boldsymbol{b}) \tag{4}$$

where $\boldsymbol{h}(\boldsymbol{x})$ represents the output value. $\boldsymbol{W}$ and $\boldsymbol{b}$ are model parameters, σ refers to the Sigmoid activation function. The

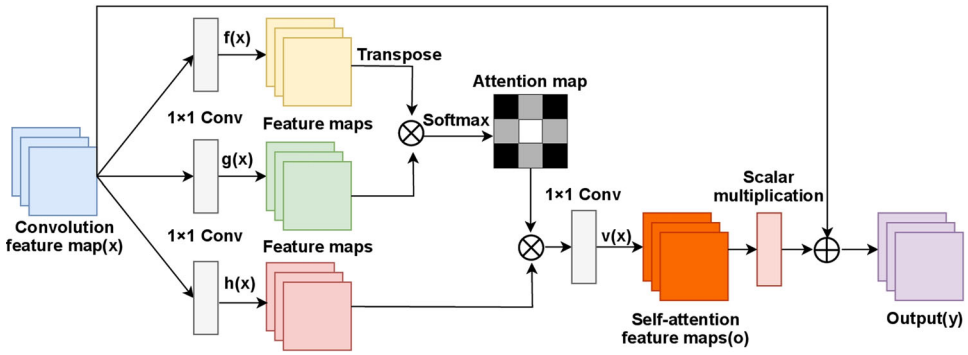


Fig. 1 Network flowchart of the self-attention mechanism layer

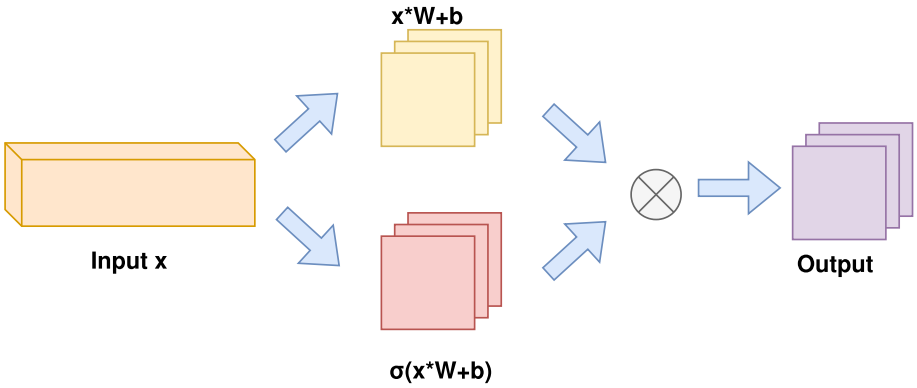


Fig. 2 GLU layer structure diagram

advantages of GLU are mainly in two aspects: On the one hand, GLU can weight data and retain the information of data in a certain proportion. It can retain important data features while discarding unimportant data features. On the other hand, gated linear elements have the ability to fit nonlinear functions while retaining linear operations, which can mitigate the problem of gradient disappearance to a certain extent [17].

The improved network not only retains the linear operation, which can accelerate the convergence speed of the network, but also does the attention mechanism operation, and filters and weights the features. In addition, there is an embedding component in the SA-GVAE proposed by us. It encodes the class label of the input image into the same information as the latent spatial dimension and inputs it to the decoder together with the output of the encoder through the multiplication layer. In this way, the SA-GVAE learns the label information and carries out supervised training. The overall network of SA-GVAE is shown in Fig. 3.

3.2 Supervise initialization and adversarial training

After the original image is input to the encoder, reparameterization is applied to convert the mean and variance learned by the encoder into data features, then the data features are fed into the decoder to output the reconstructed image. After the model is trained, the reconstructed image gradually approximates the original image. Data features of latent space tend to be normally distributed while retaining original data information, and its loss and calculation formula are shown in Eqs. 5, 6 and 7 [34].

$$\begin{aligned}(\mathfrak{g}^*, \mathfrak{h}^*) &= \arg\min KL\left(\mathfrak{q}_x(y), p(\mathfrak{y} \mid x)\right) \\ &= \arg\min \left(\mathbb{E}_{y\sim\mathfrak{q}_x}\left(\log \mathfrak{q}_x(y)\right)\right. \\ &\quad \left.-\mathbb{E}_{y\sim\mathfrak{q}_x}\left(\log \frac{\mathfrak{p}(x \mid y)\mathfrak{p}(y)}{\mathfrak{p}(x)}\right)\right) \\ &= \arg\max \left(\mathbb{E}_{y\sim\mathfrak{q}_x}(\log \mathfrak{p}(x \mid y))\right. \\ &\quad \left.-KL\left(\mathfrak{q}_x(y), \mathfrak{p}(y)\right)\right) \end{aligned} \tag{5}$$

$$z = \mu + \exp(0.5 \times \sigma) \times \alpha, \quad \alpha \sim \mathrm{U}(0, 1) \tag{6}$$

$$\mathrm{O} = z \odot e(z, \mathrm{y}) \tag{7}$$

where z is the re-parameterized variable in the latent space, the mean is μ, the standard deviation is σ, α is the normally distributed random number, and its value is uniformly distributed U(0, 1), e and y represent the embedding component and class label, respectively, and O is the output of the embedding component.

The discriminator and generator in GAN are initialized with the weight of pre-trained SA-GVAE, and a good starting point for GAN adversarial training is established. The generator is designed to have the same network topology as decoder, so the weight of pre-trained decoder can be transferred to the generator. The initialization of discriminator transfers the first few layers of encoder to it. The last layer is the dense layer matching the output dimension, which

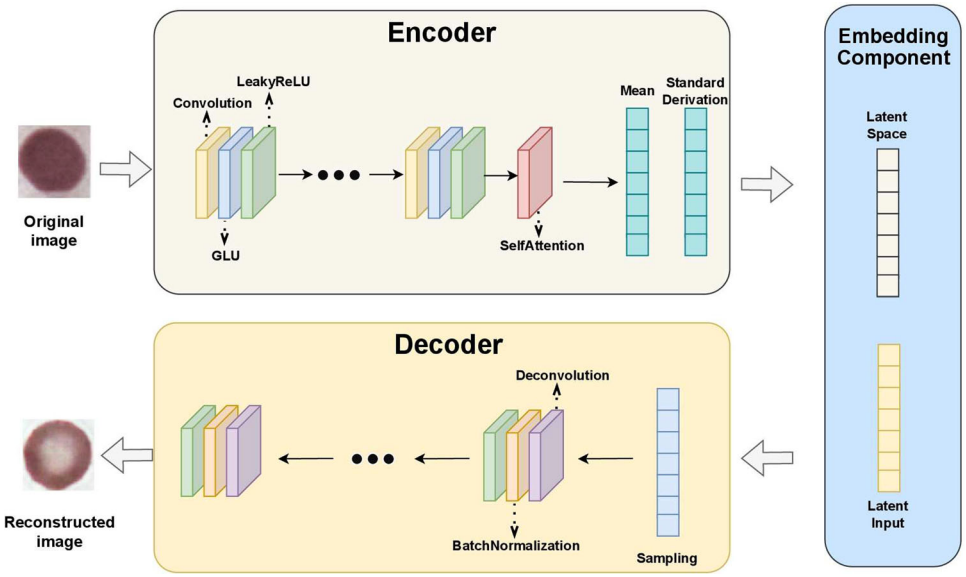


Fig. 3 Network framework of SA-GVAE

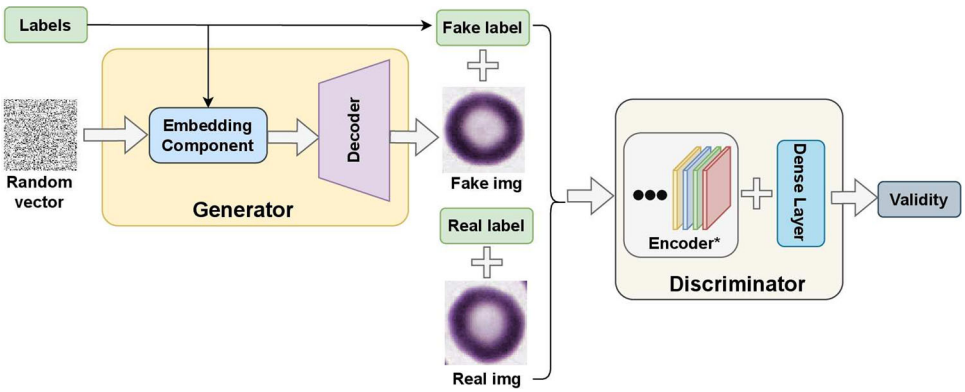


Fig. 4 SGBGAN model and adversarial training

constitutes the Pre-trained Gated Variational Autoencoder with Self-attention for Balancing generative adversarial network (SGBGAN). Its training flow chart is shown in Fig. 4, Encoder* in the figure indicates the first few layers of pre-trained encoder. After the adversarial training of SGBGAN, the generator can generate high-quality and representative samples of a specific class, and the discriminator outputs the total validity of true or false and class matching, so that the generated images can be identified as belonging to one class or false samples.

In the process of generative adversarial training, the discriminator and generator update the parameters according to the minimum and maximum adversarial setting of Eq. 8 [35].

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{8}$$

where z represents random noise input and x represents the distribution of the generated data. The loss is similar to DRAGAN [36]. The generator has only one false image loss, while the discriminator has three losses, namely false image loss, real image loss and false label loss. In order to make the discriminator converge faster, gradient penalty [37] is applied to the discriminator loss, as shown in Eq. 9.

$$GP = \lambda \mathbb{E}_{\hat{x}\sim\hat{X}} \left[\left(\left\| \nabla_{\hat{x}} D(\hat{x}) \right\|_2 - 1 \right)^2 \right] \tag{9}$$

where $\hat{x} = \alpha x_r + (1-\alpha)x_{\text{noise}}$, $\alpha \sim \text{U}(0,1)$, x_r is the real image of the input, α is a normal distribution of random numbers, obeying uniform distribution U(0, 1), λ refers to gradient punish weights, $\left\| \nabla_{\hat{x}} D(\hat{x}) \right\|_2$ is the norm of gradients. Both discriminator and generator achieve better performance by minimizing their losses during training.

3.3 Improved objective function

L2 loss minimization is applied to train the SA-GVAE network between real and reconstructed images instead of mean absolute error (MAE) [38]. Because MAE is unable to learn the class distribution of input samples, aiming at pixel to pixel similarity, and MAE is not suitable for training the supervised

VAE which is more advanced and complex [39]. The loss of SA-GVAE proposed by us consists of three parts: Kullback–Leibler (KL) divergence, cross-entropy loss and mean square error (MSE).

In order to enable the model to learn the distribution of latent space, KL divergence is used to measure the difference between the two distributions, optimizes and learns the mean and derivative of the distribution of hidden space. By minimizing the KL divergence, the decoder samples and produces better results. The other two are reconstruction losses. Cross-entropy loss is more suitable for the Bernoulli distribution [40], while MSE is more suitable for the Gaussian distribution [41]. Combining these two losses enables the model to learn more complex distribution and obtains better generative ability. Then, summing them to obtain the total loss function, the formula is shown in Eq. 10.

$$\begin{aligned} \text{Objective} &= D_{KL}(p\|q) + H(p,q) + MSE, \\ D_{KL}(p\|q) &= \sum_x p(x) \ln \frac{p(x)}{q(x)}, \\ H(p,q) &= -\sum_x p(x) \log q(x), \\ MSE &= \frac{1}{m} \sum_{i=1}^{m} \left(x_i - \hat{x}_i\right)^2 \end{aligned} \tag{10}$$

where $p(x)$ represents the distribution of real data x, $q(x)$ represents the posterior distribution, and m is the number of samples.

4 Experimental results and discussion

4.1 Datasets

We use three well-known datasets for the experiment, Fashion-MNIST and CIFAR-10, two general visual datasets, which are class-balanced datasets, each containing ten categories. And a medical small-scale blood cell dataset named Cells, which is a highly unbalanced medical image dataset, contains one majority class, namely red blood cells, and three minority classes, namely ring, schizont and trophozoite. It can be downloaded via the following link: https://storage.googleapis.com/exam-deep-learning/train.zip [14]. All image input sizes are set to 64×64.

4.2 Imbalance rate

The imbalance rate (IR) is defined as the ratio between the most majority class and the least minority class, as shown in Eq. 11. It is used to represent the difference in sample size.

$$IR = \frac{N_{\text{majority-class}}}{N_{\text{minority-class}}} \tag{11}$$

For the two balanced general visual datasets, we choose the first class as the majority class, and treat the other nine classes as the minority class. Then, we sample the minority classes to impose imbalance on Fashion-MNIST and CIFAR-10 and manually construct unbalanced datasets with imbalance rates of 5, 20, 50 and 100 for the experiment. The original imbalance rate of Cells is 52.83, and the original imbalance rate is used for the experiment.

4.3 Evaluation indicators and comparison methods

We use the Fréchet inception distance (FID) and structural similarity index measure (SSIM) to evaluate the quality of the generated image. FID is used to calculate the feature-level distance between the generated sample distribution and the real sample distribution and low FID scores correlate strongly with higher quality images. The calculation formula is shown in Eq. 12 [42].

$$FID = \left\|\mu_r - \mu_g\right\|^2 + \text{Tr}\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r \Sigma_g\right)^{1/2}\right) \tag{12}$$

where μ_r is the mean value of the real feature, μ_g is the mean value of the generated feature, Σ_r is the covariance matrix of the real feature, and Σ_g is the covariance matrix of the generated feature.

The SSIM is an index that compares the mean, variance, and covariance of the gray levels of two images. The greater the value of structural similarity index, the higher the similarity of two pictures. The calculation formula of SSIM is shown in Eq. 13 [43].

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{13}$$

where μ_x and μ_y are the mean gray values, σ_x^2 and σ_y^2 represent the gray variances, σ_{xy} represents the gray covariance between the pixel gray values x and y of the two images being compared, and c_1, c_2 are two constants to maintain stability of the denominator and numerator.

For each category in each dataset, the generated sample is compared to the test sample to calculate the corresponding FID and SSIM. DCGAN [44], BAGAN, cWGAN-GP and BAGAN-GP are used as baseline models for comparison.

4.4 Experimental setting

We use the TensorFlow 2.6 framework to implement our experiment. Detailed values of the hyperparameters are summarized in Table 2. The same setting of hyperparameters is also used in the comparison experiment. An NVIDIA-SMI Tesla V100 is used to conduct the training.

Table 2 Hyperparameter optimization for SGBGAN

Hyperparameter	Values
Learning Rate (SA-GVAE)	0.0002, 0.0005, 0.0008, 0.001
SA-GVAE Epoch	15, 30, 100,100
Adam beta1 (SA-GVAE)	0.5, 0.6, 0.7, 0.8
Learning Rate (Generator)	0.0001, 0.0002, 0.0005, 0.0008
Learning Rate (Discriminator)	0.0001, 0.0008, 0.0013, 0.0015
Gradient Penalty Weight	5
Train Ratio	5
Batch Size	128
Latent Dimension	128

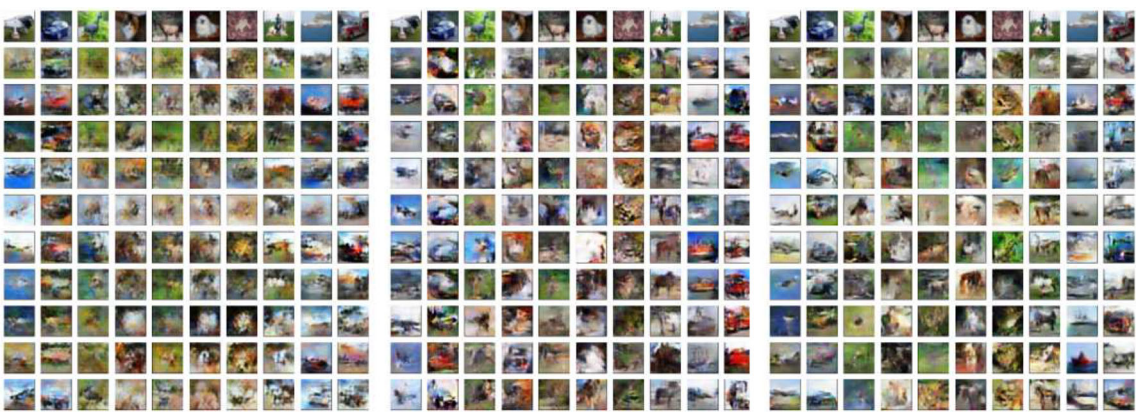
4.5 Results of general visual datasets

The unbalanced versions of Fashion-MNIST and CIFAR-10 are tested using imbalance rates of 5, 20, 50, and 100. When the imbalance rate is 100, the results of the two datasets are shown in Figs. 5 and 6. By comparing the images generated by DCGAN, BAGAN-GP and SGBGAN, it can be found that the images generated by DCGAN and BAGAN-GP are fuzzy, and most of the images lack details and textures. In addition, part of the sample generated by the DCGAN looks like unidentifiable noise, and it can be seen that most of the images generated by BAGAN-GP lack diversity and have mode collapse. The images generated by SGBGAN are more realistic and have captured the texture details of the main objects, clearly identifying the categories of each sample, and there is no pattern crash problem.

The average FID and SSIM values of one majority class and nine minority classes are shown in Table 3. It is not dif-

(a) DCGAN (b) BAGAN-GP (c) SGBGAN

Fig. 5 Image generated by Fashion-MNIST with an imbalance rate of 100



(a) DCGAN (b) BAGAN-GP (c) SGBGAN

Fig. 6 Image generated by CIFAR-10 with an imbalance rate of 100

Table 3 FID and SSIM under various unbalanced versions of Fashion-MNIST and CIFAR-10

		Fashion-MNIST				CIFAR-10			
		FID		SSIM		FID		SSIM	
IR	Model	Minority	Majority	Minority	Majority	Minority	Majority	Minority	Majority
5	DCGAN	343.35	290.33	2.71×10^{-1}	2.83×10^{-1}	482.37	385.42	5.96×10^{-2}	8.36×10^{-2}
	BAGAN	298.31	256.36	2.74×10^{-1}	2.89×10^{-1}	413.33	356.15	6.03×10^{-2}	7.75×10^{-2}
	cWGAN-GP	210.36	196.25	2.89×10^{-1}	2.91×10^{-1}	356.14	306.67	6.68×10^{-2}	8.45×10^{-2}
	BAGAN-GP	169.46	120.16	2.84×10^{-1}	$\mathbf{2.92\times10^{-1}}$	224.16	223.36	6.35×10^{-2}	8.58×10^{-2}
	SGBGAN	**163.45**	**118.04**	$\mathbf{2.93\times10^{-1}}$	2.81×10^{-1}	**210.35**	**203.35**	$\mathbf{6.94\times10^{-2}}$	$\mathbf{8.67\times10^{-2}}$
20	DCGAN	384.25	298.35	2.69×10^{-1}	2.81×10^{-1}	498.65	403.99	5.88×10^{-2}	7.65×10^{-2}
	BAGAN	303.35	263.35	2.63×10^{-1}	2.73×10^{-1}	427.58	361.17	6.19×10^{-2}	7.73×10^{-2}
	cWGAN-GP	231.35	193.65	2.75×10^{-1}	2.77×10^{-1}	369.16	336.45	6.42×10^{-2}	7.45×10^{-2}
	BAGAN-GP	176.93	126.56	2.75×10^{-1}	2.53×10^{-1}	239.46	234.38	6.53×10^{-2}	7.58×10^{-2}
	SGBGAN	**169.06**	**120.16**	$\mathbf{2.88\times10^{-1}}$	$\mathbf{2.89\times10^{-1}}$	**219.23**	**212.46**	$\mathbf{6.63\times10^{-2}}$	$\mathbf{8.23\times10^{-2}}$
50	DCGAN	392.53	323.64	2.51×10^{-1}	2.62×10^{-1}	506.40	437.65	5.63×10^{-2}	7.37×10^{-2}
	BAGAN	326.15	286.94	2.56×10^{-1}	2.65×10^{-1}	457.36	384.86	5.77×10^{-2}	7.44×10^{-2}
	cWGAN-GP	256.32	213.95	$\mathbf{2.63\times10^{-1}}$	2.66×10^{-1}	376.36	354.11	5.79×10^{-2}	7.68×10^{-2}
	BAGAN-GP	182.45	**127.46**	2.58×10^{-1}	$\mathbf{2.86\times10^{-1}}$	244.63	239.35	5.92×10^{-2}	8.01×10^{-2}
	SGBGAN	**178.32**	129.15	2.61×10^{-1}	2.81×10^{-1}	**220.32**	**213.65**	$\mathbf{6.13\times10^{-2}}$	$\mathbf{8.14\times10^{-2}}$
100	DCGAN	402.35	325.61	2.48×10^{-1}	2.53×10^{-1}	531.60	466.67	5.12×10^{-2}	6.85×10^{-2}
	BAGAN	365.32	316.36	2.51×10^{-1}	2.61×10^{-1}	488.66	416.83	6.03×10^{-2}	$\mathbf{7.75\times10^{-2}}$
	cWGAN-GP	273.21	243.76	2.56×10^{-1}	2.57×10^{-1}	387.33	363.95	6.68×10^{-2}	7.45×10^{-2}
	BAGAN-GP	180.94	128.45	2.55×10^{-1}	2.58×10^{-1}	248.16	244.62	6.35×10^{-2}	7.54×10^{-2}
	SGBGAN	**177.91**	**122.97**	$\mathbf{2.63\times10^{-1}}$	$\mathbf{2.65\times10^{-1}}$	**233.65**	**226.01**	$\mathbf{6.94\times10^{-2}}$	7.67×10^{-2}

The bold shows the best result after comparing the results produced by all models

ficult to see that SGBGAN and BAGAN-GP are superior to DCGAN, BAGAN and cWGAN-GP in the results of various imbalance rates on the two datasets, which indicates that SGBGAN and BAGAN-GP can generate high-quality images on the class-balanced datasets, and are also effective in processing unbalanced datasets. We further find that SGBGAN outperforms BAGAN-GP in almost all settings, indicating that SGBGAN is better than BAGAN-GP in processing class-imbalanced datasets. Moreover, it can be easily seen from Fig. 7 that the FID of the minority class images generated by SGBGAN is better than that of BAGAN, and the comparison is more obvious on CIFAR-10, and when the imbalance rate increases, SGBGAN is able to maintain a low FID score despite extreme imbalance rates. The same is true for the experimental results of SSIM, as shown in Fig. 8; the SSIM value of BGGAN-GP fluctuates greatly, and BAGAN-GP tends to produce unstable results. The results of SGBGAN for minority categories are all lower than those of BAGAN-GP, proving that the results generated by SGBGAN are more similar to the original images.

4.6 Medical imaging benchmark test results

Medical imaging datasets are often unbalanced in categories, and manual acquisition of real minority images is costly and time-consuming, so generative model plays an indispensable role. As a dimensionality reduction algorithm, t-Distributed Stochastic Neighbor Embedding (t-SNE) can visualize high-dimensional data and display clustering results in two-dimensional space [45]. It can be seen from the two-dimensional t-SNE diagram of encoded latent vectors (Fig. 9), the original encoder cannot learn the scattered data distribution of latent space, and the unsegmented latent vectors combine with labels to produce misleading labeled potential vectors, which misleads GAN to generate other few categories of images. However, SA-GVAE applies the embedded model to learn scattered data distribution, which makes the SGBGAN training process stable.

As shown in Fig. 10, comparing with the images generated by DCGAN, BAGAN-GP and SGBGAN, due to the lack of detail and texture capture, the quality of the samples generated by the first two models is not as good as that of SGBGAN. It can be seen that many images are covered with dark or gray color, which covers most of the shapes of cells and leads to unrecognition. Otherwise, the images generated

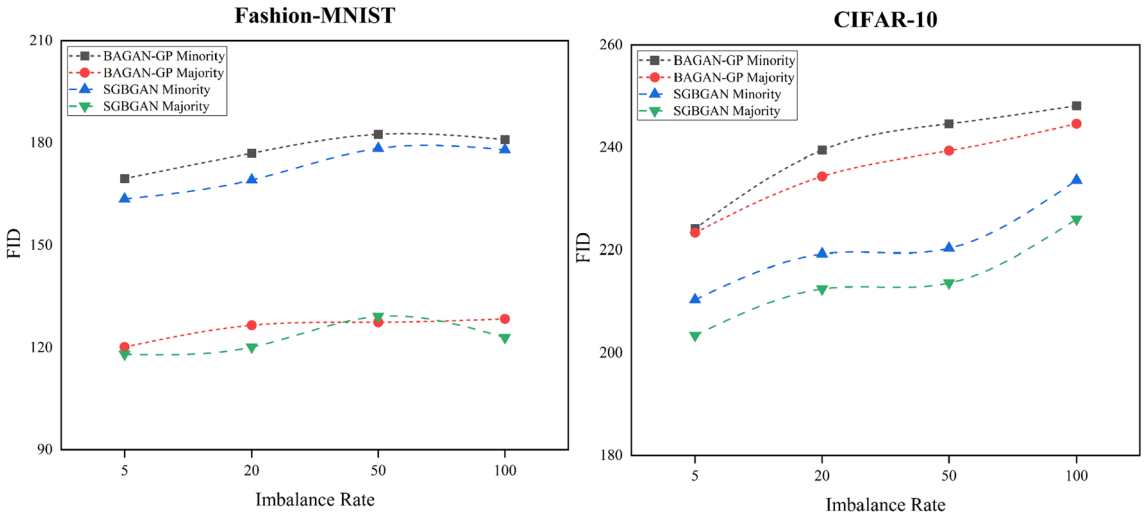


Fig. 7 FID comparison of images generated by BAGAN-GP and SGBGAN

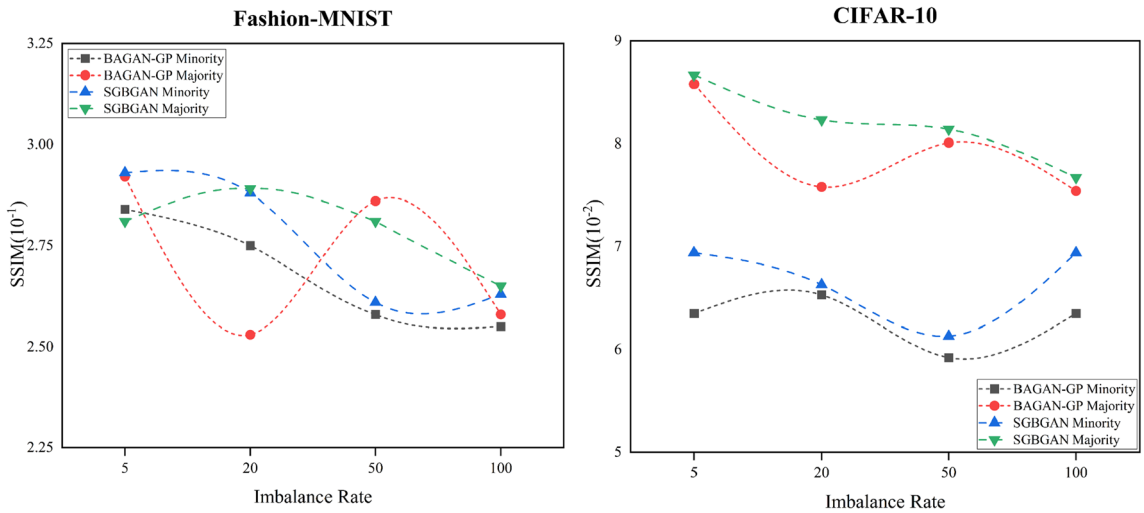


Fig. 8 SSIM comparison of images generated by BAGAN-GP and SGBGAN

by SGBGAN are brighter and more diversified with clear detail texture, small noise but no impact on recognition, and it have brighter colors and higher contrast.

The FID and SSIM scores on Cells are shown in Table 4. Obviously, we can see the influence of class-imbalanced dataset on DCGAN, BAGAN and cWGAN-GP, the minority class of experiments has worse results, but BAGAN-GP and SGBGAN have both achieved good results. This indicates that the traditional generative model cannot adapt to the highly unbalanced medical imaging dataset, and cannot solve the restrictions brought by class imbalance. In addition, comparing SGBGAN with the latest BAGAN-GP, it can be found that significant improvement has been achieved, especially for the minority of FID, and the categories of ring and trophoblast increased by 23.09% and 12.71%, respectively. The improvement of SSIM is also significant, which increases by 10.81% and 10.65% respectively for these two categories. The experimental results show that SGBGAN can generate high-quality images for the minority class when processing class-imbalanced datasets.

As shown in Fig. 11, real samples (•) and generated samples (×) are compared through the feature layer output of ResNet-50. The generated images are distributed near the real images, so they can be regarded as efficacious enhanced images. It is not difficult to find that the generated image manifolds of each category are evenly distributed around the real image manifolds, indicating that the generator of SGBGAN will not produce pattern collapse, after learning the data distribution of real samples, the generated images have diversity.

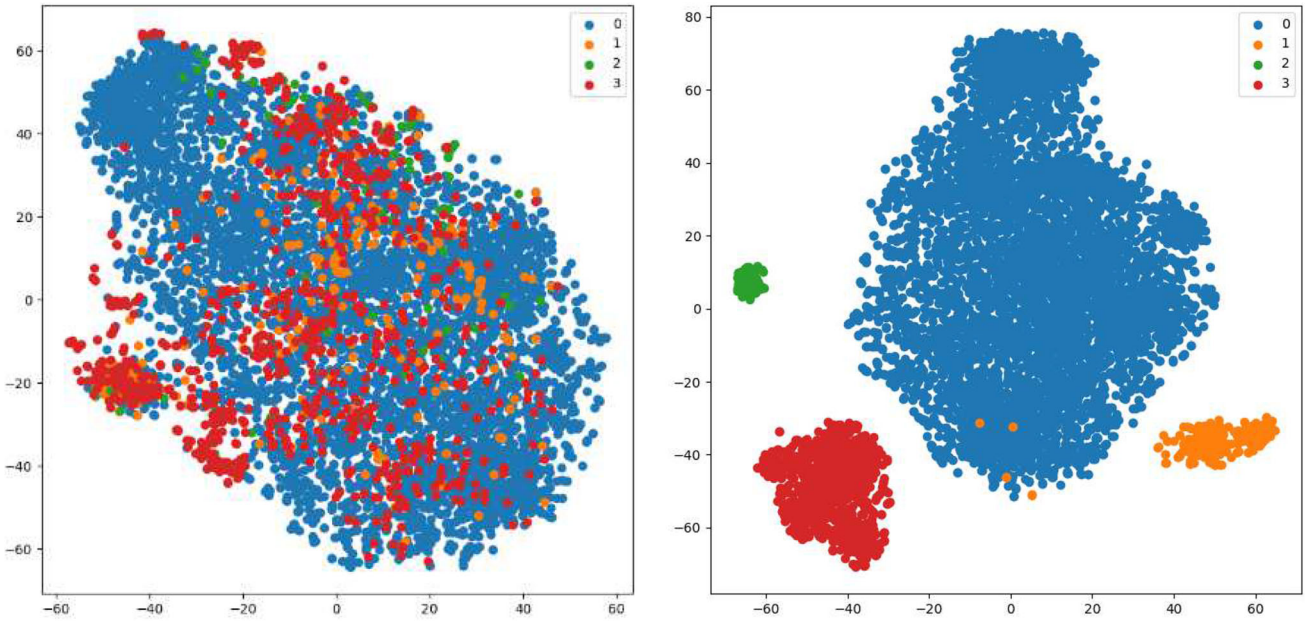


Fig. 9 Two-dimensional t-SNE diagram of encoded latent vectors

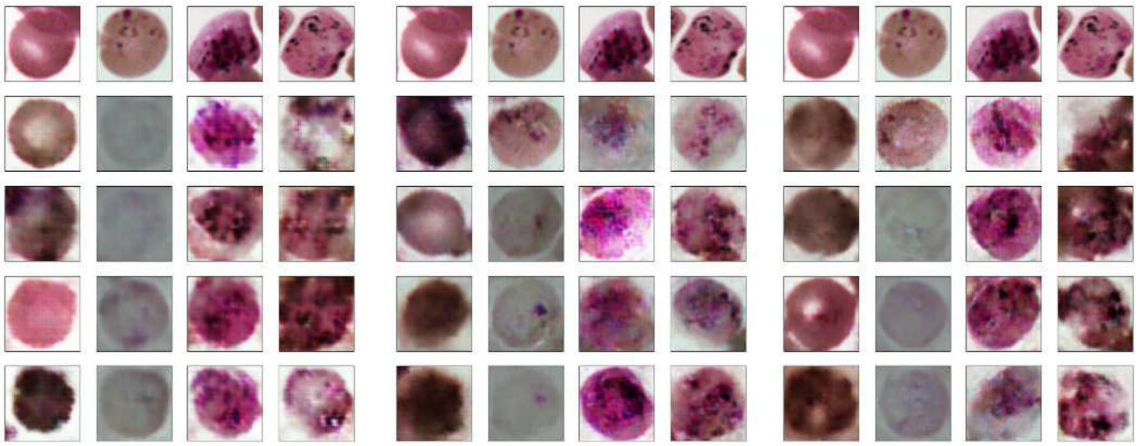


Fig. 10 Image generated by Cells with original imbalance rate

Table 4 FID and SSIM of Cells under original imbalance rate

		Cells							
		FID				SSIM			
IR	Model	Type0	Type1	Type2	Type3	Type0	Type1	Type2	Type3
52.83	DCGAN	226.99	280.32	498.33	266.13	4.66×10^{-1}	3.75×10^{-1}	3.44×10^{-1}	2.95×10^{-1}
	BAGAN	186.36	203.21	453.24	233.97	5.39×10^{-1}	4.74×10^{-1}	4.43×10^{-1}	3.65×10^{-1}
	cWGAN-GP	104.34	163.14	253.14	143.34	6.45×10^{-1}	5.54×10^{-1}	6.12×10^{-1}	5.01×10^{-1}
	BAGAN-GP	114.72	159.12	216.43	158.16	6.87×10^{-1}	5.61×10^{-1}	6.47×10^{-1}	5.54×10^{-1}
	SGBGAN	**103.80**	**122.38**	**205.68**	**137.96**	$\mathbf{7.38\times10^{-1}}$	$\mathbf{6.22\times10^{-1}}$	$\mathbf{6.96\times10^{-1}}$	$\mathbf{6.13\times10^{-1}}$
Improvement	VS BAGAN-GP	9.52%	**23.09%**	4.97%	12.71%	7.42%	**10.81%**	7.57%	10.65%

The bold shows the best result after comparing the results produced by all models

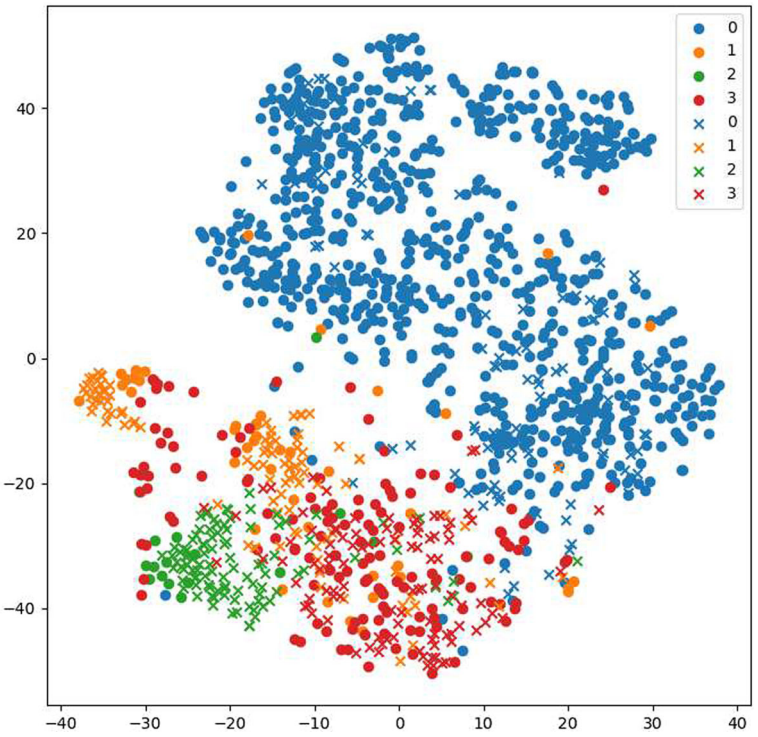


Fig. 11 Real samples (•) and generated samples (×) of ResNet-50 feature layer output

5 Conclusion

In this work, we propose a new variational autoencoder, namely SA-GVAE, which introduces self-attention mechanisms and gate linear units, and it has an embedding component to be trained in a supervisory manner. A new objective function is proposed to improve the generation and reconstruction ability of the autoencoder. By transferring pre-trained SA-GVAE weights to GAN to constitute SGBGAN, better initial guidance training is established. The training of SGBGAN follows the adversarial setting, and the loss of discriminator has gradient penalty, which can make GAN training more stable. We conducted experiments on two general visual datasets, Fashion-MNIST and CIFAR-10, and a medical imaging dataset, Cells. Imbalance is applied to the originally balanced dataset and the resulting results are compared with the four baseline models, DCGAN, BAGAN, cWGAN-GP and BAGAN-GP. The results show that SGBGAN not only can generate high-quality samples under the condition of original balance, but also can generate minority samples of higher quality under datasets with various degrees of imbalance. Even under the condition of extreme imbalance, the results are better than the four baseline models. The manifolds of the generated images prove that the results can be viewed as effectively enhanced images, and the model has learned the data distribution of the real image.

Acknowledgements The research is supported by the National Natural Science Foundation of China under Grant No. 62072468.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

Appendix A Nomenclature and abbreviations

All acronyms and full names in the paper are shown in Table 5, and unexplained variables and physical meanings are shown in Table 6.

Table 5 List of abbreviations

Abbreviation	Full term
GANs	Generative adversarial networks
BAGAN	Balancing generative adversarial networks
BAGAN-GP	BAGAN with gradient penalty
BWGAN-GP	Balanced Wasserstein generative adversarial network with gradient penalty
SA-GVAE	Gated Variational Autoencoder with Self-attention
SGBGAN	Pre-trained Gated Variational Autoencoder with Self-attention for Balancing generative adversarial network
GLU	Gated linear unit
DCGAN	Deep Convolutional Generative Adversarial Network
MAE	Mean absolute error
KL	Kullback–Leibler
MSE	Mean square error
FID	Fréchet Inception Distance
SSIM	Structure Similarity Index Measure
t-SNE	t-Distributed Stochastic Neighbor Embedding
IR	Imbalance Rate

Table 6 List of Symbols

Symbol	Variable name or physical meaning
$\beta_{j,i}$	Attention coefficient
s_{ij}	Similarity score between query i and key j
N	Total number of keys
f(x)	Query vectors, vector representations to query the inputs
g(x)	Key vectors, vector representations to store input information
h(x)	Value vectors, usually the word vectors or input representations
v(x)	Output vectors, the output vectors generated after attention calculation
$\mathfrak{g}^*$, $\mathfrak{h}^*$	Optimal parameters of the generator and discriminator
$q_x(y)$	Hypothetical posterior distribution of latent variable y by the generative network
$p(y\|x)$	True data joint distribution
$p(x\|y)$	Generative network from latent variable y to observable variable x
$p(y)$	True prior distribution of latent variable
$p(x)$	Marginal distribution of observed data
$V(D, G)$	Value function of discriminator and generator
$p_{\text{data}}(x)$	True data distribution
$p_z(z)$	Prior noise distribution
$\hat{X}$	Distribution of generated samples
$D(\hat{x})$	Discriminator's output on a generated sample
$\nabla_{\hat{x}} D(\hat{x})$	Generate the gradient of the sample against the discriminator

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 770–778 (2015)
2. Xiao, H., Rasul, K., Vollgraf, R.: Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. arXiv preprint arXiv:1708.07747 (2017)
3. Krizhevsky, A., Hinton, G., et al.: Learning multiple layers of features from tiny images (2009)
4. 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, pp. 248–255 (2009)
5. Braytee, A., Liu, W., Anaissi, A., Kennedy, P.J.: Correlated multi-label classification with incomplete label space and class imbalance. ACM Trans. Intell. Syst. Technol. (TIST) **10**, 1–26 (2019)
6. Johnson, J.M., Khoshgoftaar, T.M.: Survey on deep learning with class imbalance. J. Big Data **6**(1), 1–54 (2019)
7. Rezaei, M., Uemura, T., Näppi, J., Yoshida, H., Lippert, C., Meinel, C.: Generative synthetic adversarial network for internal bias correction and handling class imbalance problem in medical image diagnosis. In: Medical Imaging 2020: Computer-Aided Diagnosis, vol. 11314, pp. 82–89. SPIE (2020)
8. Arjovsky, M., Bottou, L.: Towards principled methods for training generative adversarial networks. arXiv preprint arXiv:1701.04862 (2017)
9. Dauphin, Y.N., Fan, A., Auli, M., Grangier, D.: Language modeling with gated convolutional networks. In: Proceedings of the 34th International Conference on Machine Learning-Volume 70, pp. 933–941 (2017)
10. Adiga, N., Pantazis, Y., Tsiaras, V., Stylianou, Y.: Speech enhancement for noise-robust speech synthesis using wasserstein gan. In: INTERSPEECH, pp. 1821–1825 (2019)
11. Zhang, H., Goodfellow, I.J., Metaxas, D.N., Odena, A.: Self-attention generative adversarial networks. arXiv:1805.08318 (2018)
12. Gurunlu, B., Ozturk, S.: Efficient approach for block-based copy-move forgery detection. In: Smart Trends in Computing and Communications: Proceedings of SmartCom 2021, pp. 167–174. Springer (2022)
13. Mariani, G., Scheidegger, F., Istrate, R., Bekas, C., Malossi, A.C.I.: Bagan: Data augmentation with balancing gan. arXiv:1803.09655 (2018)
14. Huang, G., Jafari, A.H.: Enhanced balancing gan: minority-class image generation. Neural Comput. Appl. **35**, 5145–5154 (2023)
15. Zhang, M., Xiao, T.Z., Paige, B., Barber, D.: Improving vae-based representation learning. arXiv preprint arXiv:2205.14539 (2022)
16. Taghanaki, S.A., Havaei, M., Lamb, A., Sanghi, A., Danielyan, A., Custis, T.: Jigsaw-vae: Towards balancing features in variational autoencoders. arXiv:2005.05496 (2020)
17. Zheng, Y., Ma, Y., Tian, C.: Tmrn-glu: A transformer-based automatic classification recognition network improved by gate linear unit. Electronics **11**(10), 1554 (2022)
18. Li, Z., Jin, Y., Li, Y., Lin, Z., Wang, S.: Imbalanced adversarial learning for weather image generation and classification. In: 2018 14th IEEE International Conference on Signal Processing (ICSP), pp. 1093–1097 (2018)
19. Shoohi, L.M., Saud, J.H.: Dcgan for handling imbalanced malaria dataset based on over-sampling technique and using cnn. Med. Leg. Update **20**, 1079–1085 (2020)
20. Wang, Q., Zhou, X., Wang, C., Liu, Z., Huang, J., Zhou, Y., Li, C., Zhuang, H., Cheng, J.-Z.: Wgan-based synthetic minority over-sampling technique: Improving semantic fine-grained classi-

fication for lung nodules in ct images. IEEE Access **7**, 18450–18463 (2019)
21. Rai, H., Shukla, N.: Unpaired image-to-image translation using cycle-consistent adversarial networks (2018)
22. Balasubramanian, R., Sowmya, V., Gopalakrishnan, E.A., Menon, V.K., Variyar, V.V.S., Soman, K.P.: Analysis of adversarial based augmentation for diabetic retinopathy disease grading. In: 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1–5 (2020)
23. Waheed, A., Goyal, M., Gupta, D., Khanna, A., Al-turjman, F., Pinheiro, P.R.: Covidgan: Data augmentation using auxiliary classifier gan for improved covid-19 detection. IEEE Access **8**, 91916–91923 (2020)
24. Sampath, V., Maurtua, I., Martín, J.J.A., Gutierrez, A.: A survey on generative adversarial networks for imbalance problems in computer vision tasks. J. Big Data **8** (2020)
25. Chen, J., Tam, D., Raffel, C., Bansal, M., Yang, D.: An empirical survey of data augmentation for limited data learning in nlp. Trans. Assoc. Comput. Linguist. **11**, 191–211 (2023)
26. Xu, M., Yoon, S., Fuentes, A., Park, D.S.: A comprehensive survey of image augmentation techniques for deep learning. Pattern Recognit. 109347 (2023)
27. Zheng, M., Li, T., Zhu, R., Tang, Y., Tang, M., Lin, L., Ma, Z.: Conditional wasserstein generative adversarial network-gradient penalty-based approach to alleviating imbalanced data classification. Inf. Sci. **512**, 1009–1023 (2020)
28. Dai, W., Li, D., Tang, D., Wang, H., Peng, Y.: Deep learning approach for defective spot welds classification using small and class-imbalanced datasets. Neurocomputing **477**, 46–60 (2022)
29. Xu, M., Chen, Y., Wang, Y., Wang, D., Liu, Z., Zhang, L.: Bwgan-gp: An eeg data generation method for class imbalance problem in rsvp tasks. IEEE Trans. Neural Syst. Rehabil. Eng. **30**, 251–263 (2022)
30. Ding, N., Zhang, G., Zhang, L., Shen, Z., Yin, L., Zhou, S., Deng, Y.: Engineering an ai-based forward-reverse platform for the design of cross-ribosome binding sites of a transcription factor biosensor. Comput. Struct. Biotechnol. J. **21**, 2929–2939 (2023)
31. Snoussi, R., Youssef, H.: Vae-based latent representations learning for botnet detection in iot networks. J. Netw. Syst. Manag. **31**(1), 4 (2023)
32. Zhao, H., Jia, J., Koltun, V.: Exploring self-attention for image recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10076–10085 (2020)
33. Dauphin, Y., Fan, A., Auli, M., Grangier, D.: Language modeling with gated convolutional networks. In: International Conference on Machine Learning (2016)
34. Yao, Y., Wangr, X.L., Ma, Y., Fang, H., Wei, J., Chen, L., Anaissi, A., Braytee, A.: Conditional variational autoencoder with balanced pre-training for generative adversarial networks. In: 2022 IEEE 9th International Conference on Data Science and Advanced Analytics (DSAA), pp. 1–10 (2022)
35. Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A.C., Bengio, Y.: Generative adversarial nets. In: NIPS (2014)
36. Kodali, N., Hays, J., Abernethy, J.D., Kira, Z.: On convergence and stability of gans. arXiv:Artificial Intelligence (2018)
37. Arjovsky, M., Chintala, S., Bottou, L.: Wasserstein generative adversarial networks. In: International Conference on Machine Learning (2017)
38. Willmott, C.J., Matsuura, K.: Advantages of the mean absolute error (mae) over the root mean square error (rmse) in assessing average model performance. Clim. Res. **30**(1), 79–82 (2005)
39. Kingma, D.P., Welling, M.: Auto-encoding variational bayes. CoRR arXiv:1312.6114 (2013)
40. Yeung, M., Sala, E., Schönlieb, C.-B., Rundo, L.: Unified focal loss: Generalising dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Comput. Med. Imaging Graph. **95**, 102026 (2022)
41. Sara, U., Akter, M., Uddin, M.S.: Image quality assessment through fsim, ssim, mse and psnr-a comparative study. J. Comput. Commun. **7**(3), 8–18 (2019)
42. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: NIPS (2017)
43. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
44. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR arXiv:1511.06434 (2015)
45. Wattenberg, M., Viégas, F., Johnson, I.: How to use t-sne effectively. Distill **1**(10), 2 (2016)

Qian Wan received the B.Sc. degree in electrical engineering and automation from University of Jinan, Jinan, China, in 2021. He is pursuing the M.Sc. degree with the College of Control Science and Engineering, China University of Petroleum (East China), Qingdao 266580, Shandong Province, China. His research interests include computer vision, deep learning and image processing.

Wenhui Guo received the B.Sc. degree in applied mathematics from Xinzhou Teachers University, Xinzhou, China, in 2017, and received the M.Sc. degree in applied mathematics, China Jiliang University, Hangzhou, China. She is currently pursuing the Ph.D. degree in Control Science and Engineering, China University of Petroleum, Qingdao, China. Her research interests include deep learning, image processing and EEG processing.

Yanjiang Wang is currently a full Professor in the College of Control Science and Engineering, China University of Petroleum (East China), China. He received the M.S. degree in communication and electronic system from Beijing University of Aeronautics and Astronautics, Beijing, China, in 1989 and the Ph.D. degree in signal and information processing from Beijing Jiaotong University, Beijing, China, in 2001. From 2013 to 2014, he was a visiting scholar in the department of psychological and brain sciences, Indiana University, Bloomington, IN, USA. His current research interests include bio-inspired pattern recognition, cognitive memory modeling and human brain connectivity. He has authored or co-authored more than 200 papers in top journals and prestigious conferences.