



Article

Generative Adversarial Network for Class-Conditional Data Augmentation

Jeongmin Lee , Younkyoung Yoon and Junseok Kwon * 

School of Computer Science and Engineering, Chung-Ang University, Seoul 156-756, Korea; minlee077@naver.com (J.L.); amyss0907@gmail.com (Y.Y.)

* Correspondence: jskwon@cau.ac.kr; Tel.: +82-10-8693-7455

Received: 2 November 2020; Accepted: 24 November 2020; Published: 26 November 2020



Abstract: We propose a novel generative adversarial network for class-conditional data augmentation (i.e., GANDA) to mitigate data imbalance problems in image classification tasks. The proposed GANDA generates minority class data by exploiting majority class information to enhance the classification accuracy of minority classes. For stable GAN training, we introduce a new denoising autoencoder initialization with explicit class conditioning in the latent space, which enables the generation of definite samples. The generated samples are visually realistic and have a high resolution. Experimental results demonstrate that the proposed GANDA can considerably improve classification accuracy, especially when datasets are highly imbalanced on standard benchmark datasets (i.e., MNIST and CelebA). Our generated samples can be easily used to train conventional classifiers to enhance their classification accuracy.

Keywords: generative adversarial network; data augmentation; image classification

1. Introduction

In imbalanced datasets, training data are not uniformly distributed over the different classes. Despite the imbalance, these datasets are widely used in many image classification problems because it is nontrivial to gather data evenly for all classes. In many cases, we inevitably have to train conventional classifiers using imbalanced datasets, which induces highly biased or imprecise classification results. This problem occurs due to lack of consideration of the class ratio and the balance of the dataset.

Data augmentation can help mitigate this data imbalance problem by generating new synthetic data for imbalanced classes and improving the balance between classes. However, conventional data augmentation methods (e.g., mirroring, rotation, and geometric transformation) have several potential problems. For example, translation and random cropping may run the risk of changing image labels. Flip or rotation may disrupt objects' orientation-related features. In particular, for domain-dependent applications such as medical image analysis, there is more severe bias between potentially generated data (augment data) and training data than translational or positional variances.

Unlike geometric transformations, generative adversarial networks (GANs) [1] can improve the general balance among classes while remaining less affected by the application domain and dataset characteristics. GANs generate minority class images that can be used to restore the balance of classes in datasets. Because GANs can generate realistic samples that are uncovered by original datasets, they have shown outstanding performance in terms of data augmentation. However, despite their capabilities, GANs have difficulty in stably optimizing their objective functions. In addition, if minority data are very scarce, conventional GANs still face generalization problems. To overcome the aforementioned shortcomings of GANs for imbalanced datasets, Balanced GAN (BAGAN) [2] has been proposed, which is specialized in imbalanced datasets. Based on autoencoder initialization and class conditioning in the latent space method, BAGAN attempts to describe the minority class distribution

using the data, which is jointly learned by majority and minority classes. However, BAGAN also has several issues that need to be addressed. First, BAGAN cannot explicitly impose the class condition in the latent space, which yields unintended class samples, especially when different class regions are significantly overlapped with each other in the latent space. Second, the aforementioned problem makes it difficult to use BAGAN to train deep neural networks, and BAGAN fails to optimize the network.

In this paper, we solve the data imbalance problem for image classification tasks and propose a novel data augmentation method based on GANs (i.e., generative adversarial network for class-conditional data augmentation (GANDA)), which can produce minority class data for accurate classification by leveraging majority class information effectively. To achieve stable learning in the GAN framework, we present a new denoising autoencoder initialization. Using this technique, we can leverage a more accurate representation of each class relatively in the generator latent space. Moreover, we introduce a conditional one-hot code, which can be used to generate definite samples. Experimentally, our proposed method can accurately generate minority samples from various imbalanced datasets where the BAGAN method has failed to do so.

Our generated samples are visually realistic and have a high resolution ($128 \times 128 \times 3$). When our generated samples are used for training conventional classifiers, the proposed GANDA significantly enhances the classification accuracy of existing classifiers and shows state-of-the-art performance given imbalanced datasets in various experiments. Figures 1 and 2 show the differences of the proposed GANDA from conventional methods. The main contributions of this work are as follows.

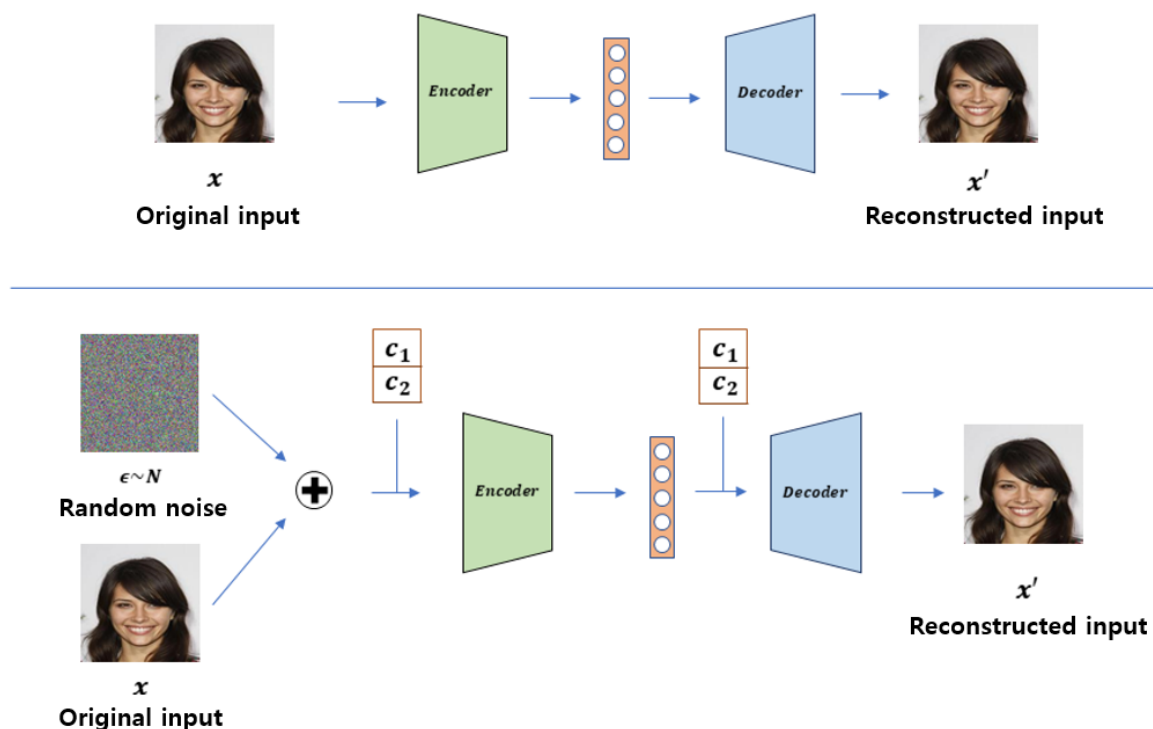


Figure 1. Conceptual comparison between balanced generative adversarial network (BAGAN) autoencoder and the proposed generative adversarial network for class-conditional data augmentation (GANDA) autoencoder. Random noises are inserted into an input image to diversify output images. The conditional one-hot code is inserted into the proposed encoder and decoder to generate definite samples.

- We present a novel data augmentation method based on GANs. Our proposed method can generate minority class data accurately in imbalanced datasets.
- For stable GAN training, we present a new denoising autoencoder initialization technique with explicit class conditioning in the latent space.

- We conduct various experiments showing underlying problems in conventional methodologies. We experimentally show that majority class data can help generate minority class data and considerably enhance its classification accuracy.

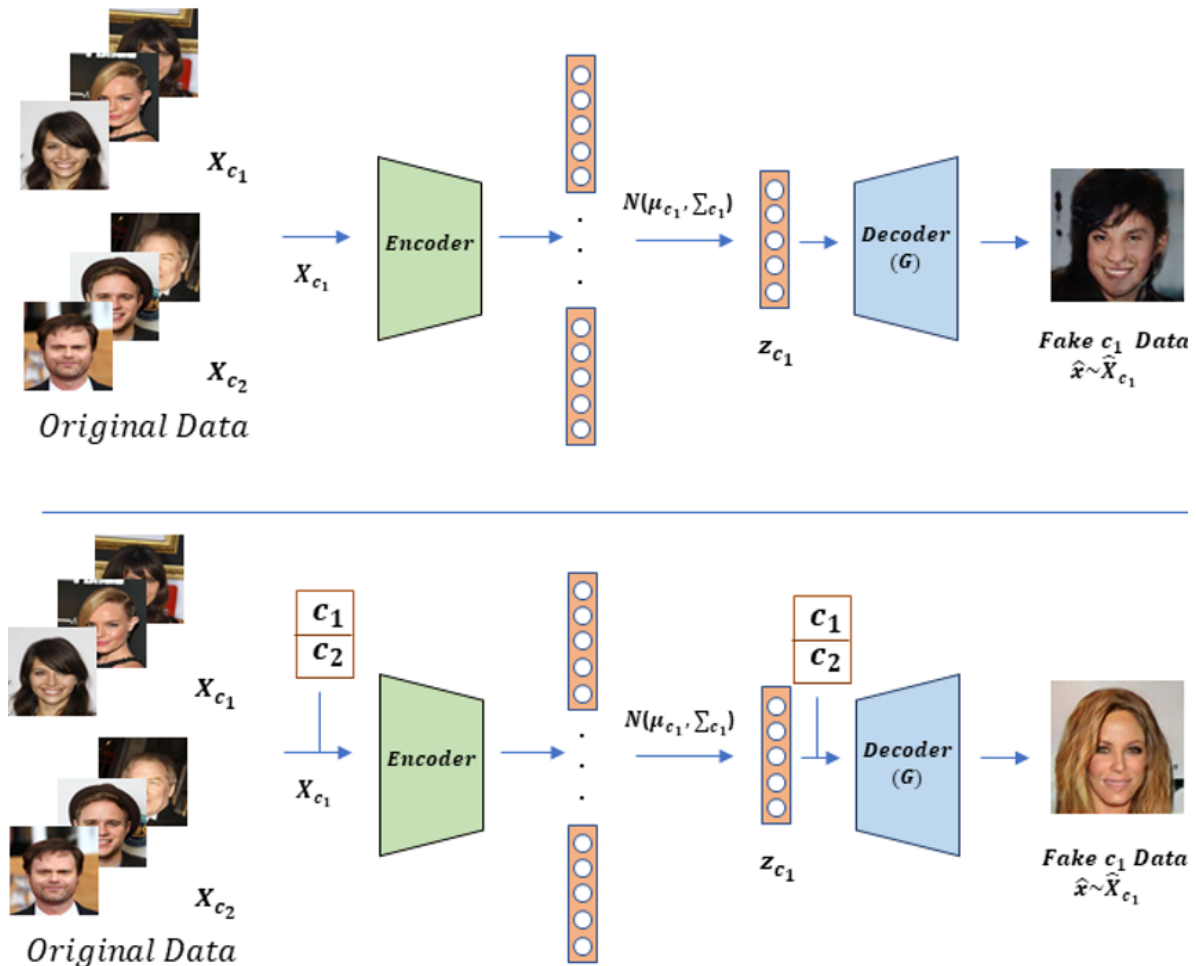


Figure 2. Conceptual comparison between BAPAN generator and the proposed GANDA generator.

2. Related Work

2.1. Data Augmentation Methods

There are several approaches for data augmentation [3]. In traditional ways to augment data, images are geometrically transformed or distorted using rotation, scaling, white-balancing, or sharpening [4–6]. A relatively recent and powerful tool has been to use generative adversarial networks such as ACGAN [7] and BAPAN [2]. ACGAN can adapt to select the target class to be generated, but does not consider minority classes in image classification.

In contrast to ACGAN, our method directly targets the imbalanced dataset problem in image classification tasks. Our idea mainly stems from BAPAN, which can jointly solve data generation and image classification problems. However, the proposed GANDA produces more accurate classification results and induces fewer overlaps among different classes in the latent space using our explicit class conditioning. Please note that it is nontrivial to perform class conditioning in the latent space for data augmentation because there are many very similar classes, and class overlaps can occur easily.

2.2. Methods for Imbalanced Datasets

There are two representative ways that can be used to solve the imbalanced data problems. One is to resample the data via either oversampling or undersampling [8–13]. A synthetic minority oversampling technique (SMOTE) [14] combined oversampling and undersampling. For oversampling, SMOTE applied data augmentation to minority samples to mitigate the overfitting issue. However, oversampling is susceptible to overfitting problems, whereas undersampling usually results in the loss of valuable information in majority samples. The second method is cost-sensitive learning, which aims to avoid the aforementioned issues by assigning different costs for the misclassification of the minority class [15–20]. For example, dynamic curriculum learning [21], or active learning [22], has been proposed to tackle the problem of dataset imbalance. Some approaches [23,24] view the class-imbalance problem within the meta-learning framework. Max-marginal learning [25] directly enforced the max-margin constraints. Similar to this method, our method also attempts to delineate the space between the classes more prominently. However, it is difficult to design proper cost functions in different problem settings or environments in the aforementioned cost-sensitive learning. To solve the imbalanced data problem, M2m [26] generated minority samples using majority samples.

In contrast to these methods, our proposed GANDA uses autoencoders and GANs to generate minority class data. Please note that DOPING [27] originally aimed to detect anomaly events, but it can be used for augmenting minority class data, because oversampling anomaly samples at the boundary of the latent distributions is equivalent to augmenting minority class data. While DOPING also adopted autoencoders and GANs, the latent space is not tailored by class conditioning for more accurate oversampling, which is not similar to our proposed method.

2.3. GAN-Based Methods

After the seminal work proposed by Ian Goodfellow et al. [1], GANs have been actively researched to generate realistic fake data. For example, many studies have used GANs for imitation [28–30]. BAGAN [2] addressed imbalance problems by coupling autoencoders and GANs. It generated minority class data and jointly performed classification tasks to drive the generation of desired minority classes. BAGAN produced more accurate classification results compared to ACGAN [7], which separately handled data generation and image classification tasks. However, SCDL-GAN [31] shows that BAGAN cannot avoid mode collapse problems, in which its generator produces limited varieties of samples. SCDL-GAN produced impressive results by using a Wasserstein-based autoencoder in the representation of the class distributions.

In contrast, the proposed GANDA adopts explicit class conditioning to avoid mode collapse problems and to more accurately generate the minority classes.

3. The Proposed Method

3.1. Difficulties in Autoencoder Initialization

We observe that conventional autoencoder initialization with implicit class conditioning in the latent space can cause the latent space to train unstably [2]. In addition, conventional generators can generate wrong samples even though the training converges successfully. We first examine why the aforementioned problems occur through the following experiments. First and foremost, we carry out t-SNE visualization for class-conditional codes in the latent space, which are determined by the autoencoder initialization.

Figure 3a illustrates the t-SNE of the conventional autoencoder initialization using the MNIST datasets. As shown in the figure, autoencoder initialization can maximize inter-class distance while minimizing intra-class distance. In addition, it causes visually similar classes to become distributed close to each other in the latent space. For example, the “9” (cyan) and “4” (red) classes share visually similar characteristics; thus, they are positioned relatively closer than other classes. Using autoencoder initialization, minority classes can also be recognized in the latent space if we examine the relative

distance of other majority class data in the latent space. However, class-conditional noises learned by autoencoder initialization can cause unintended effects in the latent space if there is no explicit class conditioning. As shown in Figure 3a, many overlapped areas appear for visually similar classes, which potentially produces wrong class samples. For example, “4” and “9” are two different but very adjacent classes in the latent space. Thus, conventional approaches can frequently generate “9” when sampling “4”.

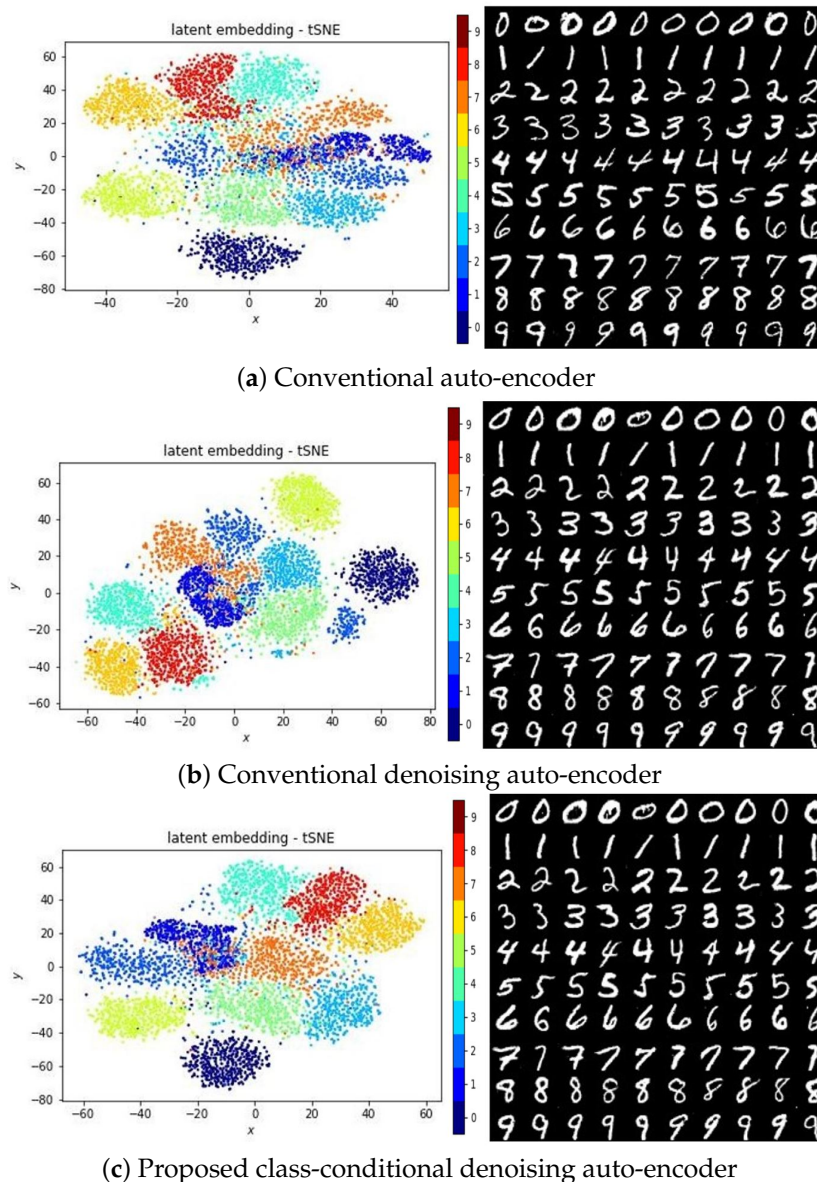


Figure 3. Difficulty in conventional autoencoder initialization.

To solve this problem, a denoising autoencoder [32] has been proposed to widen the gap between different classes in the latent space by training robust features for different classes. Figure 3b shows the t-SNE results using the denoising autoencoder. We qualitatively observe that denoising autoencoders broaden the inter-class gap in the latent space and shrink the intra-class gap compared with original autoencoders. However, the denoising autoencoder could not fully solve the inter-class overlapping problem as there still exists overlapped areas among different classes. Thus, in terms of quantitative measures, denoising autoencoders are similar to original autoencoders. In addition, if we handle high-resolution images in a low dimensional latent space, the aforementioned problem still remains and even worsens, as explained in the next section. Please note that Figure 3c illustrates t-SNE,

which visualizes the proposed GANDA encoder representation. Although our GANDA uses a class-conditional one-hot code, it effectively disentangles different classes in the latent space, as shown in Figure 3c.

3.2. Difficulties in High-Resolution Data Generation

To show difficulties in high-resolution data generation, we conduct experiments using BAGAN with autoencoder initialization on the CelebA dataset ($128 \times 128 \times 3$) and perform a simple binary classification of males and females.

Figure 4 shows generated samples for females using class-conditional noise. As shown in the figure, conventional methods typically generated correct samples but also created unintended samples. Some samples even wrongly belong to the other class (male). We argue that this problem occurs because class conditioning is not explicitly imposed for adjacent areas between classes in the latent space.

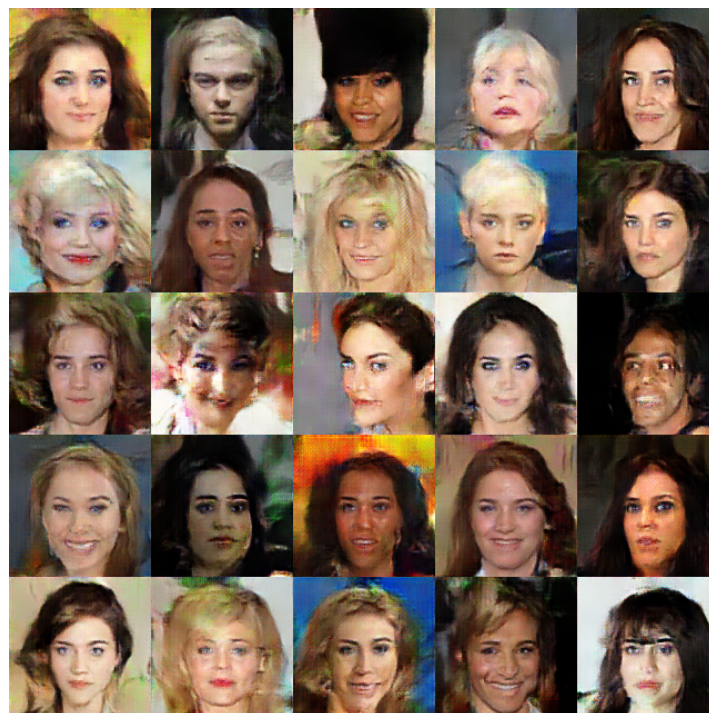


Figure 4. Difficulty in high-resolution data generation.

Moreover, this problem becomes even worse for minority classes, because we cannot accurately represent their latent spaces due to the lack of class data. Figure 4 shows generated data with high resolution using the CelebA dataset. As shown in the figure, conventional methods generated many unintended class samples.

The problems of conventional approaches (e.g., BAGAN) can be summarized as follows. First, the conventional autoencoding process does not explicitly regard the imbalanced datasets. Therefore, it can incorrectly learn the latent space of the minority class compared to the majority class. Second, conventional approaches utilize the majority class data to generate minority class samples using implicit class conditioning in the latent space. Therefore, unintended class samples can be generated. Finally, existing approaches experimentally produce inaccurate results in that samples from other (majority) classes are used to generate images that belong to a certain (minority) class.

3.3. Class-Conditional GAN-Based DA

As shown in the experiments in Figure 3, autoencoder initialization can broaden the inter-class gap in the latent space and shrunken the intra-class gap compared with original autoencoders.

Thus, autoencoder initialization helps GANs to stably train imbalanced datasets. However, if samples are generated only with class-conditional noises, the training becomes unstable due to the possible overlaps between classes in the latent space. To settle this problem, we propose a variant of the GAN architecture called GANDA. In GANDA, minority classes are effectively generated using data from both majority and minority classes in imbalanced datasets to restore the balance of datasets. To generate a specific class sample, the proposed method explicitly conditions the class label to our generator. At the same time, class-conditional noises, which contain relative distances in the latent space with other classes, is used to leverage other class information. As the training proceeds, our method generates realistic minority classes and simultaneously improves the classification accuracy of both majority and minority classes.

The proposed method consists of two main steps: conditional denoising autoencoder initialization and adversarial learning. We describe each step in detail, as follows.

- **Conditional denoising autoencoder initialization:** We first add noise ϵ to an input image $x \in \mathbb{R}^d$, which can diversify output images. Subsequently, we concatenate the noisy input image with a one-hot label $y = \{c_1, c_2, \dots, c_n\}$ and feed it into the encoder $\phi \in \mathbb{R}^{d \times p}$, in which class-aware latent features z are extracted as follows,

$$z = \phi(x + \epsilon, y). \quad (1)$$

The estimated latent feature is concatenated with the same one-hot label y once more to explicitly guide the generation process. The concatenated feature is then fed into the decoder $\psi \in \mathbb{R}^{p \times d}$, which results in x' as follows,

$$x' = \psi(z, y). \quad (2)$$

We train the proposed conditional denoising autoencoder using the reconstruction loss \mathcal{L}_{recon} :

$$\mathcal{L}_{recon}(x, x') = \|x - x'\|_2. \quad (3)$$

To optimize the objective function in (3), we adopt the l_2 norm. Please note that the trained autoencoder describes the distribution of the latent codes of the classes. Thus, we initialize a part of the discriminator, D_e , and generator, G , using the parameters of encoder ϕ and decoder ψ , respectively.

- **Adversarial learning:** After conditional denoising autoencoder initialization, we train generator G and discriminator D via adversarial training. Before the adversarial learning proceeds, we use the encoder ϕ to determine the multinomial distribution of the class latent code in the training data. Then, the generator is trained to produce a fake class sample \hat{x}_y by selecting the latent code z_y , which is extracted from the aforementioned multinomial distribution determined by ϕ and is concatenated with the corresponding one-hot label y .

$$\hat{x}_y = G(z_y | y). \quad (4)$$

The discriminator classifies the generated sample as belonging to one of n classes or as being fake.

$$D(x|y) = P_D(y = i|x, y) = \frac{\exp(l_i)}{\sum_{j=1}^{k+1} \exp(l_j)}, \quad (5)$$

where l_j denotes the j -th output of the discriminator. Please note that the discriminator can have $(n + 1)$ outputs in total, where n is the number of class labels, and 1 is for the fake label. The objective function for adversarial learning is as follows,

$$\begin{aligned} \min_G \max_D V(D, G) = & \mathbb{E}_{x \sim p_{data}(x)} [\log D(x|y)] \\ & + \mathbb{E}_{z_y \sim p_z(z|y)} [\log(1 - D(G(z_y|y)))]. \end{aligned} \quad (6)$$

The advantages of the proposed method can be summarized as follows. Our class conditioning makes different classes to be well separated in the latent space even when minority class data are scarce. Thus, class conditioning enables accurate data augmentation and classification of the minority class data. Moreover, class conditioning can allow our method to safely use the features, which are shared by different classes. For example, facial images of males and females include common features like eyes, ears, mouth, and nose, where two classes could be differentiated by hair length and make-up. Then, we can use the shared features from majority data (i.e., male) to generate minority data (i.e., female), while accurately separating these two types of data in the latent space. In Section 4.3, we verify that our method generates minority class data accurately using majority class data. Please note that we can use the class information without additional cost in supervised learning.

4. Experiments

4.1. Implementation Details

Network architecture details: The proposed method consists of two main components for conditional denoising autoencoder initialization and adversarial learning. For stable learning, we add the spectral normalization layer [33] to the encoder (i.e., part of the discriminator) and imposed the Lipschitz constraint. In addition, the decoder adopts transposed convolution layer, of which the structure is same as that of the generator. For more details about the network architecture, please refer to Table 1.

Table 1. Network architectures for the CelebA dataset ($128 \times 128 \times 3$). The slopes of all lReLU functions were set to 0.1.

Name	Input Size	Output Size
Encoder	$128 \times 128 \times 3, 2 \times 1$	100×1
Conv + SN + lReLU	$128 \times 128 \times 3$	$64 \times 64 \times 32$
Conv + SN + lReLU	$64 \times 64 \times 32$	$64 \times 64 \times 64$
Conv + SN + lReLU	$64 \times 64 \times 64$	$32 \times 32 \times 128$
Conv + SN + lReLU	$32 \times 32 \times 128$	$16 \times 16 \times 256$
Conv + SN + lReLU	$16 \times 16 \times 256$	$8 \times 8 \times 256$
Conv + SN + lReLU	$8 \times 8 \times 256$	$8 \times 8 \times 256$
Conv + SN + lReLU	$8 \times 8 \times 256$	$4 \times 4 \times 256$
Conv + SN + lReLU	$4 \times 4 \times 256$	$4 \times 4 \times 256$
Flatten	$4 \times 4 \times 256$	4096×1
Concat	$4096 \times 1, 2 \times 1$	4098×1
Dense	4098×1	100×1
Decoder (Generator)	$100 \times 1, 2 \times 1$	$128 \times 128 \times 3$
Concat	$100 \times 1, 2 \times 1$	102×1
Dense + ReLU	102×1	1024×1
Dense + ReLU	1024×1	8192×1
Tconv + ReLU	$8 \times 8 \times 128$	$16 \times 16 \times 128$
Tconv + ReLU	$16 \times 16 \times 128$	$32 \times 32 \times 128$
Tconv + ReLU	$32 \times 32 \times 128$	$64 \times 64 \times 64$
Tconv + ReLU	$64 \times 64 \times 64$	$128 \times 128 \times 32$
Tconv + Tanh	$128 \times 128 \times 32$	$128 \times 128 \times 3$
Discriminator	$128 \times 128 \times 3, 2 \times 1$	11×1
Encoder (partial)	$128 \times 128 \times 3, 2 \times 1$	$8 \times 8 \times 256$
flatten	$8 \times 8 \times 256$	16384×1
Dense + Softmax	16384×1	11×1

Hyperparameters: The proposed network is trained with a batch size of 32. We use the ADAM optimizer with a fixed learning rate of 0.00005, $\beta_1 = 0.5$, and $\beta_2 = 0.9999$. For conditional

denoising autoencoding, we train the encoder ϕ and decoder ψ with 150 epochs by using the L2 loss. For denoising, we use a standard normal function $N(0, I)$ to make noise and combine the noise with an input image using a weight of 0.5 to make a noisy image. If the standard deviation is high, we can make heavy and diverse images with noises. We initialize the weights of the discriminator (generator) with the encoder (decoder). Subsequently, we train the discriminator and generator via adversarial learning using the sparse categorical cross-entropy loss function.

Datasets: We conduct experiments using the MNIST [34] and CelebA [35] datasets, where MNIST contains around 50 K handwritten images of 28×28 in size with 10 different classes. For CelebA, we randomly extract 10 K images of males and females, respectively, and resized them to 128×128 resolutions. We conduct the ablation test (i.e., in-depth analysis through evaluating the proposed method in a component-wise manner) by changing the degree of class imbalance. For this experiment, we use the MNIST and CelebA datasets. In the case of the MNIST dataset, we consider class 0 as a minority class and removed 60%, 80%, 90%, 95%, and 97.5% of the training images for class 0. In the case of the CelebA dataset, we treat class *female* as a minority class and removed 60%, 70%, 80%, and 90% of the training images.

Evaluation metrics: To evaluate the image generation performance, two metrics are used, namely, inception score (IS) [36] and Fréchet inception distance (FID) [37]. To evaluate the classification performance, we use the average classification accuracy and validation score [38].

4.2. Ablation Study

To provide in-depth analysis and insights on the proposed GANDA, we conduct several ablation studies.

Data augmentation from majority to minority classes: We experimentally show that the majority classes can help augment minority classes in the proposed GANDA framework. For this experiment, we examine latent code interpolation, which explains how the latent codes work in the GANDA framework. As shown in the generated images in Figure 5, the conditional one-hot code c_y for $y = \{male, female\}$ represents the main feature that determines a specific class, whereas the conditional noise z_y expresses the detailed features (e.g., hair length, clothing, and skin color). Therefore, the experimental results demonstrate that the information obtained by learning majority classes through the mapping of latent code z can also be used to express shared features of the minority classes. Figure 6 shows images generated by autoencoding and the proposed method for the CIFAR10 dataset. As shown in Figure 6, our method qualitative outperforms conventional autoencoding approaches in terms of diversity of samples.

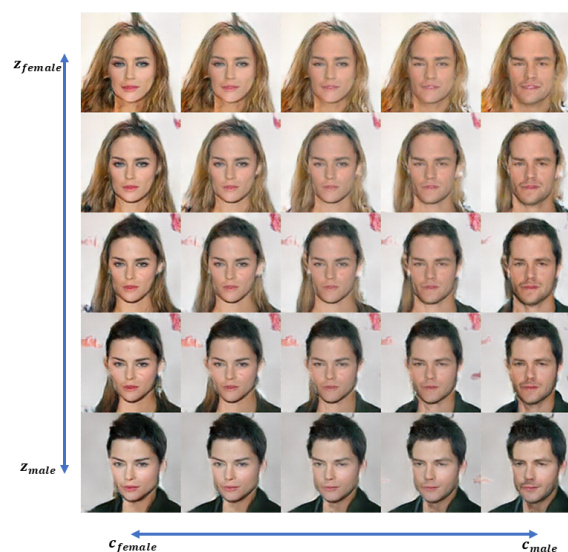


Figure 5. Latent code interpolation.

Degree of class imbalance: We verify the effectiveness of the proposed GANDA when there is a very small amount of training data in a minority class. Table 2 shows the IS and FID scores when we increase the degree of class imbalance by removing more training data from the minority class (from 60%, 80%, and 90%).

As shown in Table 2, even when we delete 90% of the training data from the minority class, the proposed GANDA outperforms the baseline BAGAN in terms of IS and FID. Thus, our proposed GANDA can augment minority class data accurately, which can be used for object classification.

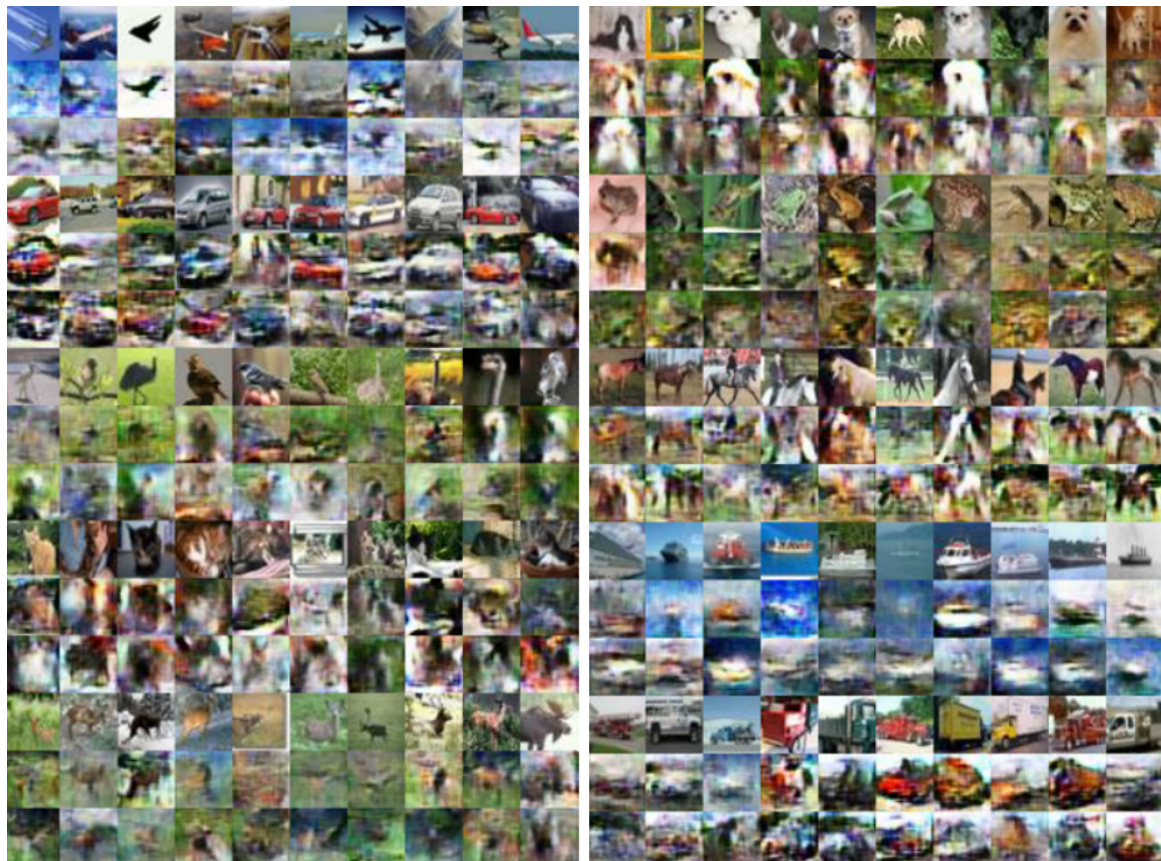


Figure 6. Data augmentation using the CIFAR-10 dataset. For each block with three rows, the first, second, and third rows contain real images, generated images using autoencoding, and generated images using the proposed method, respectively.

Table 2. Comparison of the proposed GANDA with BAGAN on the CelebA dataset in terms of IS and FID (IS: higher is better, FID: lower is better).

Method (Removal Ratio)	IS \uparrow	FID \downarrow
CelebA (real)	2.79 ± 0.09	11.53
BAGAN (0.6)	1.89 ± 0.02	79.79
BAGAN (0.7)	1.79 ± 0.03	82.20
BAGAN (0.8)	1.78 ± 0.03	138.97
BAGAN (0.9)	1.83 ± 0.02	167.62
GANDA (0.6)	2.18 ± 0.05	48.89
GANDA (0.7)	1.93 ± 0.03	65.32
GANDA (0.8)	1.93 ± 0.02	71.45
GANDA (0.9)	1.84 ± 0.02	91.86

4.3. Data Augmentation Comparison

FID and IS: Table 2 quantitatively compares the proposed GANDA with other state-of-the-art methods. Although the data imbalance worsened, the proposed GANDA generates better quality samples than the baseline BAGAN.

High resolution: Figure 7 qualitatively compares the proposed GANDA with other state-of-the-art methods [2] using the CelebA [35] dataset. The proposed GANDA produces more realistic high-resolution images than BAGAN. The noise z obtained from class-conditioning in the latent space can contain the shared features (e.g., hair length, clothing, and skin color). Thus, samples using latent noise for the female samples can be used to generate males with long hair $G(z_{female}|c_{male})$, while samples using latent noise for the male samples can be applied to produce short-haired females $G(z_{male}|c_{female})$. In addition, skin color, dress, and background are irrelevant to gender. Thus, the details do not change significantly, although the class changes.



Figure 7. Qualitative comparison for high-resolution image generation using the CelebA dataset. (a) Samples produced by BAGAN for the minority class (female). There are some unintended class (male) samples. (b) Samples produced by GANDA for the minority class (female). All samples represent intended class (female) appearance. (c) Samples synthesized by varying the label code through the GANDA framework. The first two rows show samples by fixing $z_{male} \sim N(\mu_{male}, \sigma_{male})$ and changing the one-hot label code c . The following third and fourth rows compare samples by fixing $z_{female} \sim N(\mu_{female}, \sigma_{female})$ and changing the one-hot label code c .

4.4. Data Classification Comparison

Validation-score: We evaluate the classification performance using the validation score (V-Score) [38]. The V-Score can measure the clustering accuracy. To calculate the V-Score, we need to compute two terms, which are homogeneity and completeness. On one hand, homogeneity v_h determines if each cluster has only the same label. On the other hand, completeness v_c measures if all data points of the same class label belong to the same cluster, which can be numerically evaluated as follows,

$$\begin{aligned} v_h &= \begin{cases} 1, & \text{if } H(C, K) = 0, \\ 1 - \frac{H(C|K)}{H(C)}, & \text{otherwise.} \end{cases} \\ v_c &= \begin{cases} 1, & \text{if } H(K, C) = 0, \\ 1 - \frac{H(K|C)}{H(K)}, & \text{otherwise.} \end{cases} \end{aligned} \quad (7)$$

where $C = \{c_1, \dots, c_n\}$ denotes a set of classes, and $K = \{k_1, \dots, k_m\}$ denotes a set of clusters. In (7), we normalize the conditional entropy $H(C|K)$ and $H(K|C)$ using $H(C)$ and $H(K)$, respectively, to remove class size dependencies.

Then, the V-Score \mathcal{V}_β is the weighted mean of v_c and v_h :

$$\mathcal{V}_\beta = \frac{(1 + \beta)v_h v_c}{\beta v_h + v_c}, \quad (8)$$

where parameter β can be adjusted to favor either homogeneity or completeness. In our experiment, we set β to 1, giving the same weight to both metrics. The V-score results are obtained with the proposed conditional denoising autoencoder initialization only and not with adversarial training. We apply the V-Score to the latent space obtained by the proposed conditional autoencoder and compared the latent space of our GANDA to that of BAGAN. As shown in Table 3, our method outperforms BAGAN. The performance improvements on V-Score can be obtained owing to explicit class conditioning in the latent space. For this experiment, we use the k-means algorithm to cluster the latent space and MNIST as the testing dataset, which consists of 10 classes. Please note that the number of class labels, the number of classes, the size of the data, and the clustering algorithm are independent of each other. Thus, the V-Score can accurately evaluate the quantitative classification performance.

Table 3. Comparison of the proposed GANDA with BAGAN in terms of V-Score.

	V-Score (K-Means)
GANDA (conditional denoising autoencoder initialization)	0.779
BAGAN (denoising autoencoder initialization)	0.739

Classification score: Tables 4 and 5 show the effects of using the generated samples as augmented data for classification tasks. We conduct this experiment by changing the removal ratio for the minority class. For the MNIST [34] dataset, we compare the existing methods (i.e., Vanilla GAN [1], ACGAN [7], and BAGAN [2]) with the proposed GANDA. Our GANDA outperforms state-of-the-art methods in various removal ratios. The CelebA [35] dataset consists of very high-resolution images, where the proposed GANDA also outperforms state-of-the-art methods. The classification accuracy in Tables 4 and 5 are not empirically affected by slight changes in the specific network architecture in Table 1.

Table 4. The average accuracy (%) of the minority class achieved by the ResNet-18 classifier trained with the augmented MNIST dataset whose balance is restored after removing a portion of the minority class images.

	60	80	90	95	97.5
Plain	99.13	98.87	98.62	96.51	95.4
Vanilla GAN [1]	98.96	98.92	98.35	96.64	95.12
ACGAN [7]	99.21	98.73	98.43	96.72	95.96
BAGAN [2]	99.38	98.87	98.67	97.75	96.2
GANDA (ours)	99.79	99.48	99.18	97.63	96.42

Table 5. The average accuracy (%) of the minority class achieved by the ResNet-18 classifier trained with an augmented subset of the CelebA dataset whose balance is restored after removing a portion of the minority class images.

	60	70	80	90
Plain	92.52	91.54	89.24	83.94
BAGAN	93.55	90.33	88.49	82.73
GANDA (ours)	94.59	93.67	90.79	85.73

5. Conclusions

In this paper, we proposed a novel generative adversarial network for the image classification using class conditional data augmentation (i.e., GANDA). The proposed GANDA effectively restores data imbalance in the GAN framework. For this, we presented a denoising autoencoder initialization technique with explicit class conditioning in the latent space, which provides a good initial point for GANs while effectively utilizing the information learned from majority class data to generate minority class data. We demonstrated the effectiveness of proposed method on the classification tasks using imbalanced datasets. The proposed GANDA outperforms other state-of-the-art methods.

Author Contributions: Conceptualization, J.L. and J.K.; validation, J.L. and Y.Y.; writing—original draft preparation, J.L.; writing—review and editing, Y.Y. and J.K.; supervision, J.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (No. 2020M3C1C2A01080885).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Nets. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 8–13 December 2014.
2. Mariani, G.; Scheidegger, F.; Istrate, R.; Bekas, C.; Malossi, A.C.I. BAGAN: Data Augmentation with Balancing GAN. *arXiv* **2018**, arXiv:1803.09655.
3. Mikołajczyk, A.; Grochowski, M. Data augmentation for improving deep learning in image classification problem. In Proceedings of the 2018 International Interdisciplinary PhD Workshop (IIPhDW), Swinouście, Poland, 9–12 May 2018.
4. Galdran, A.; Alvarez-Gila, A.; Meyer, M.I.; Saratxaga, C.L.; Araújo, T.; Garrote, E.; Aresta, G.; Costa, P.; Mendonça, A.M.; Campilho, A. Data-Driven Color Augmentation Techniques for Deep Skin Image Analysis. *arXiv* **2017**, arXiv:1703.03702, 2017.

5. Kwasigroch, A.; Mikołajczyk, A.; Grochowski, M. Deep convolutional neural networks as a decision support tool in medical problems—Malignant melanoma case study. In *Trends in Advanced Intelligent Control, Optimization and Automation*; Mitkowski, W., Kacprzyk, J., Oprze dkievicz, K., Skruch, P., Eds.; Springer: Berlin/Heidelberg, Germany, 2017.
6. Okafor, E.; Schomaker, L.; Wiering, M.A. An analysis of rotation matrix and colour constancy data augmentation in classifying images of animals. *J. Inf. Telecommun.* **2018**, *2*, 465–491. [[CrossRef](#)]
7. Odena, A.; Olah, C.; Shlens, J. Conditional Image Synthesis With Auxiliary Classifier GANs. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 19–24 June 2016.
8. Drummond, C.; Holte, R. C4.5, Class imbalance, and cost sensitivity: Why under-sampling beats oversampling. In Proceedings of the Workshop on Learning from Imbalanced Datasets II, Washington, DC, USA, 21–24 August 2003.
9. Estabrooks, A.; Jo, T.; Japkowicz, N. A Multiple Resampling Method for Learning from Imbalanced Data Sets. *Comput. Intell.* **2004**, *20*, 18–36. [[CrossRef](#)]
10. Han, H.; Wang, W.Y.; Mao, B.H. *Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning*; Advances in Intelligent Computing; Springer: Berlin/Heidelberg, Germany, 2005.
11. He, H.; Garcia, E.A. Learning from Imbalanced Data. *IEEE Trans. Knowl. Data Eng.* **2009**, *21*, 1263–1284.
12. Maciejewski, T.; Stefanowski, J. Local neighbourhood extension of SMOTE for mining imbalanced data. In Proceedings of the 2011 IEEE Symposium on Computational Intelligence and Data Mining (CIDM), Paris, France, 11–15 April 2011.
13. Oquab, M.; Bottou, L.; Laptev, I.; Sivic, J. Learning and Transferring Mid-Level Image Representations Using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Columbus, OH, USA, 23–28 June 2014.
14. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [[CrossRef](#)]
15. Tang, Y.; Zhang, Y.; Chawla, N.V.; Krasser, S. SVMs Modeling for Highly Imbalanced Classification. *IEEE Trans. Syst. Man, Cybern. Part B* **2009**, *39*, 281–288. [[CrossRef](#)] [[PubMed](#)]
16. Thai-Nghe, N.; Gantner, Z.; Schmidt-Thieme, L. Cost-sensitive learning methods for imbalanced data. In Proceedings of the The 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010.
17. Yang, C.Y.; Yang, J.S.; Wang, J.J. Margin calibration in SVM class-imbalanced learning. *Neurocomputing* **2009**, *73*, 397–411. [[CrossRef](#)]
18. Zadrozny, B.; Langford, J.; Abe, N. Cost-sensitive learning by cost-proportionate example weighting. In Proceedings of the Third IEEE International Conference on Data Mining (ICDM), Melbourne, FL, USA, 19–22 December 2003.
19. Zhou, Z.H.; Liu, X.Y. Training cost-sensitive neural networks with methods addressing the class imbalance problem. *IEEE Trans. Knowl. Data Eng.* **2006**, *18*, 63–77. [[CrossRef](#)]
20. Ting, K.M. A Comparative Study of Cost-Sensitive Boosting Algorithms. In Proceedings of the 17th International Conference on Machine Learning (ICML), Stanford, CA, USA, 29 June–2 July 2000.
21. Wang, Y.; Gan, W.; Yang, J.; Wu, W.; Yan, J. Dynamic Curriculum Learning for Imbalanced Data Classification. In Proceedings of the IEEE International Conference on Computer Vision (CVPR), Long Beach, CA, USA, 16–20 June 2019.
22. Aggarwal, U.; Popescu, A.; Hudelot, C. Active Learning for Imbalanced Datasets. In Proceedings of the IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass Village, CO, USA, 2–5 March 2020.
23. Liu, Z.; Miao, Z.; Zhan, X.; Wang, J.; Gong, B.; Yu, S.X. Large-Scale Long-Tailed Recognition in an Open World. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 16–20 June 2019.
24. Wang, Y.X.; Ramanan, D.; Hebert, M. Learning to Model the Tail. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
25. Hayat, M.; Khan, S.; Zamir, S.W.; Shen, J.; Shao, L. Gaussian Affinity for Max-Margin Class Imbalanced Learning. In Proceedings of the IEEE International Conference on Computer Vision (CVPR), Long Beach, CA, USA, 16–20 June 2019.

26. Kim, J.; Jeong, J.; Shin, J. M2m: Imbalanced Classification via Major-to-Minor Translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 14–19 June 2020.
27. Lim, S.K.; Loo, Y.; Tran, N.T.; Cheung, N.M.; Roig, G.; Elovici, Y. DOPING: Generative Data Augmentation for Unsupervised Anomaly Detection with GAN. In Proceedings of the IEEE International Conference on Data Mining (ICDM), Singapore, 17–20 November 2018.
28. Douzas, G.; Bao, F. Effective data generation for imbalanced learning using conditional generative adversarial networks. *Expert Syst. Appl.* **2018**, *91*, 464–471. [[CrossRef](#)]
29. Li, Z.; Jin, Y.; Li, Y.; Lin, Z.; Wang, S. Imbalanced Adversarial Learning for Weather Image Generation and Classification. In Proceedings of the 14th IEEE International Conference on Signal Processing (ICSP), Beijing, China, 12–16 August 2018.
30. Mullick, S.S.; Datta, S.; Das, S. Generative Adversarial Minority Oversampling. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 12 October–2 November 2019.
31. Cai, Z.; Wang, X.; Zhou, M.; Xu, J.; Jing, L. Supervised Class Distribution Learning for GANs-Based Imbalanced Classification. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Seoul, Korea, 12 October–2 November 2019.
32. Vincent, P.; Larochelle, H.; Lajoie, I.; Bengio, Y.; Manzagol, P.A. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *J. Mach. Learn. Res.* **2010**, *11*, 3371–3408.
33. Miyato, T.; Kataoka, T.; Koyama, M.; Yoshida, Y. Spectral Normalization for Generative Adversarial Networks. *arXiv*, **2018**, arXiv:1802.05957, 2018.
34. LeCun, Y.; Cortes, C.; Burges, C. MNIST Handwritten Digit Database. ATT Labs. 2010, Volume 2. Available online: <http://yann.lecun.com/exdb/mnist> (accessed on 5 July 2020) .
35. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Face Attributes in the Wild. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015.
36. Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; Chen, X.; Chen, X. Improved Techniques for Training GANs. In Proceedings of the Advances in Neural Information Processing Systems (NIPS), Barcelona, Spain, 5–10 December 2016.
37. 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 Proceedings of the Advances in Neural Information Processing Systems (NIPS), Long Beach, CA, USA, 4–9 December 2017.
38. Rosenberg, A.; Hirschberg, J. V-Measure: A Conditional Entropy-Based External Cluster Evaluation Measure. In Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic, 28–30 June 2007.

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).