

Improving GANs for Long-Tailed Data through Group Spectral Regularization

Harsh Rangwani¹, Naman Jaswani¹, Tejan Karmali^{1,2},
Varun Jampani², and R. Venkatesh Babu¹

¹ Indian Institute of Science, Bengaluru, India

² Google Research

Abstract. Deep long-tailed learning aims to train useful deep networks on practical, real-world imbalanced distributions, wherein most labels of the tail classes are associated with a few samples. There has been a large body of work to train discriminative models for visual recognition on long-tailed distribution. In contrast, we aim to train conditional Generative Adversarial Networks, a class of image generation models on long-tailed distributions. We find that similar to recognition, state-of-the-art methods for image generation also suffer from performance degradation on tail classes. The performance degradation is mainly due to class-specific mode collapse for tail classes, which we observe to be correlated with the spectral explosion of the conditioning parameter matrix. We propose a novel group Spectral Regularizer (gSR) that prevents the spectral explosion alleviating mode collapse, which results in diverse and plausible image generation even for tail classes. We find that gSR effectively combines with existing augmentation and regularization techniques, leading to state-of-the-art image generation performance on long-tailed data. Extensive experiments demonstrate the efficacy of our regularizer on long-tailed datasets with different degrees of imbalance. Project Page: <https://sites.google.com/view/gsr-eccv22>.

1 Introduction

Generative Adversarial Networks (GAN) [7] are consistently at the forefront of generative models for image distributions, also being used for diverse applications like image-to-image translation [23], super resolution [20] etc. One of the classic applications of GAN is class specific image generation, by conditioning on the class label y . The generated images in ideal case should associate to class label y , be of high quality and exhibit diversity. The conditioning is usually achieved with conditional Batch Normalization (cBN) [5] layers which induce class-specific (y) features at each layer of the generator. The additional class conditioning information enables GAN models like the state-of-the-art (SOTA) BigGAN [2] to generate diverse images, in comparison to unconditional models [13].

Recent works [45] demonstrate that performance of models like BigGAN deteriorates from mode collapse when limited training data is presented. The differentiable data augmentation approaches [14, 35, 45] attempt to mitigate this



Fig. 1: Regularizing GANs on long-tailed training data. (left) Images generated from BigGAN trained on long-tailed CIFAR-10. (right) FID scores vs. training steps. The proposed gSR regularizer prevents mode collapse, for the tail classes [2, 36, 45].

degradation by enriching the training data through augmentations. On the other hand, model based regularization techniques like LeCam [36] are also proposed to prevent the degradation of image quality in such cases.

Although these methods lead to effective increase in image generation quality, they are designed to increase performance when trained on balanced datasets (*i.e.* even distribution of samples across classes). We find that the SOTA methods like BigGAN (LeCam) with augmentation, which are designed for limited data, also suffer from the phenomenon of class-specific mode collapse when trained on long-tailed datasets. By *class-specific mode collapse*, we refer to deteriorating quality of generated images for tail classes, as shown in Fig. 1.

In this work we aim to investigate the cause of the class-specific mode collapse that is observed in the generated images of tail classes. We find that the class-specific mode collapse is correlated with spectral explosion (*i.e.* sudden increase in spectral norms, ref. Fig. 2) of the corresponding class-specific (cBN) parameters (when grouped into a matrix, described in Sec. 3.3). To prevent this spectral explosion, we introduce a novel class-specific *group Spectral Regularizer* (gSR), which constrains the spectral norm of class-specific cBN parameters. Although there are multiple spectral [37, 42] regularization (and normalization [26]) techniques used in deep learning literature, all of them are specific to model weights W , whereas our regularizer focuses on cBN parameters. We, through our analysis, show that our proposed gSR leads to reduced correlation among tail classes' cBN parameters, effectively mitigating the issue of class-specific mode collapse.

We extensively test our regularizer's effectiveness by combining it with popular SNGAN [26] and BigGAN [2] architectures. It also complements discriminator specific SOTA regularizer (LeCam+DiffAug [36]), as it's combination with gSR ensures improved quality of image generation even across tail classes (Fig. 1). In summary, we make the following contributions:

- We first report the phenomenon of class-specific mode collapse, which is observed when cGANs are trained on long-tailed imbalance datasets. We

find that spectral norm explosion of class-specific cBN parameters correlates with its mode collapse.

- We find that even existing techniques for limited data [15, 22, 36] are unable to prevent class-specific collapse. Hence, we propose a novel group Spectral Regularizer (gSR) which helps in alleviating such collapse.
- Combining gSR with existing SOTA GANs with regularization leads to large average relative improvement (of $\sim 25\%$ in FID) for image generation on 5 different long-tailed dataset configurations.

2 Related Work

Generative Adversarial Networks: Generative Adversarial Networks [7] are a combination of Generator G and Discriminator D aimed at learning a generative model for a distribution. GANs have enabled learning of models for complex distributions like high resolution images etc. One of the inflection point for success was the invention of Spectral Normalization (SN) for GANs (SNGAN) which enabled GANs to scale to datasets like ImageNet [6] (1000 classes). The GAN training was further scaled by BigGAN [2] which demonstrated successful high resolution image generation, using a deep ResNet network.

Regularization: Several regularization techniques [16, 21, 23, 44, 46, 48] are developed to alleviate the problem of mode collapse in GANs like Gradient Penalty [8], Spectral Normalization [26] etc. These include LeCam [36] and Differentiable Augmentations [15, 45] which are the regularization techniques specifically designed to prevent mode collapse in limited data scenarios. Commonality among majority of these techniques are that they (a) designed for the data which is balanced across classes, and (b) focus on discriminator network D . In this work, we aim to regularize the cBN parameters in generator G , which makes our regularizer complementary to earlier works.

Long-Tailed Imbalanced Data: Long-tailed imbalance is a form of distribution in which majority of the data samples are present in head classes and the occurrence of per class data samples decreases exponentially as we move towards tail classes (Fig. 1(left)). This family of distribution represents natural distribution for species' population [10], objects [38] etc. As these distributions are natural, a lot of work has been done to learn discriminative models (*i.e.* classifiers) [3, 4, 12, 25, 41, 47] which work across all classes, despite training data following a long-tailed distribution. However, even though there has been a lot of interest, there are still only a handful works which aim to learn generative models for long-tailed distribution. Mullick *et al.* [28] developed GAMO which aims to learn how to oversample using a generative model. Class Balancing GAN (CBGAN) with a Classifier in the Loop [31] is the only work which aims to learn a GAN to generate good quality images across classes (in a long-tailed distribution). However, their model is an unconditional GAN which requires a trained classifier to guide the GAN. The requirement of such a classifier can be restrictive. In this work we aim to develop conditional GANs which use class labels, does not require external classifier and generate good quality images (even for tail classes) when trained using long-tailed training data.

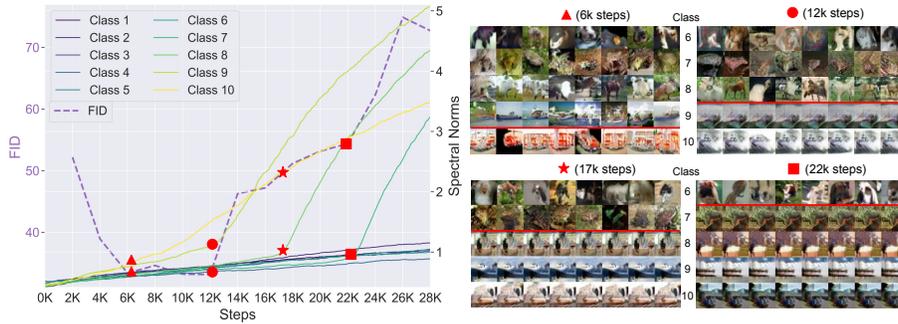


Fig. 2: Correlation between class-specific mode collapse and spectral explosion. (left) FID/Spectral Norms of class-specific gain parameter of conditional Batch-Norm layer on CIFAR-10. Symbols on plot indicate that FID score’s increase correlates with onset of spectral explosion on 4 tail classes respectively. (right) Images generated for tail classes at these train steps reveals corresponding class-specific mode collapse.

3 Approach

We start by describing conditional Generative Adversarial Networks (Sec. 3.1) and the associated class-specific mode collapse (Sec. 3.2). Following that we introduce our regularizer formulation, and explain the decorrelation among features caused by gSR that mitigates the mode collapse for tail classes (Sec. 3.3).

3.1 Conditional Generative Adversarial Networks

Generative Adversarial Networks (GAN) are a combination of two networks the generator G and discriminator D . The discriminator’s goal is to classify images from training data distribution (P_r) as real and the images generated through G as fake. In a conditional GAN (cGAN), the generator and discriminator are also enriched with the information about the class label y associated with the image $\mathbf{x} \in \mathbb{R}^{3 \times H \times W}$. The conditioning information y helps the cGAN in generating diverse images of superior quality, in comparison to unconditional GAN. The cGAN objectives can be described as:

$$\begin{aligned} \max_D V(D) &= \mathbb{E}_{x \sim P_r} [f_D D(x|y)] + \mathbb{E}_{z \sim P_z} [f_G (1 - D(G(z|y)))] \\ \min_G \mathcal{L}_G &= \mathbb{E}_{z \sim P_z} [f_G (1 - D(G(z|y)))] \end{aligned} \quad (1)$$

where p_z is the prior distribution of latents, f_D , f_G , and g_G refer to the mapping functions from which different formulations of GAN can be derived (ref. [22]). The generator $G(z|y)$ generates images corresponding to the class label y . In earlier works the conditioning information y was concatenated with the noise vector z , however recently conditioning each layer using cBN layer has shown improved performance [27]. For a feature $\mathbf{x}_y^l \in \mathbb{R}^d$ conditioned on class y (out of

K classes) corresponding to layer l , the transformation can be described as:

$$\hat{\mathbf{x}}_{\mathbf{y}}^l = \frac{\mathbf{x}_{\mathbf{y}}^l - \mu_{\mathbf{B}}^l}{\sqrt{\sigma_{\mathbf{B}}^l{}^2 + \epsilon}} \rightarrow \gamma_{\mathbf{y}}^l \hat{\mathbf{x}}_{\mathbf{y}}^l + \beta_{\mathbf{y}}^l \quad (2)$$

The $\mu_{\mathbf{B}}^l$ and $\sigma_{\mathbf{B}}^l{}^2$ are the mean and the variance of the batch respectively. The $\gamma_{\mathbf{y}}^l \in \mathbb{R}^d$ and the $\beta_{\mathbf{y}}^l \in \mathbb{R}^d$ are the cBN parameters which enable generation of the image for specific class y . We focus on the behaviour of these parameters in the subsequent sections.

3.2 Class-Specific Mode Collapse

Due to widespread use of conditional GANs [2, 27], it is important that these models are able to learn across various kinds of training data distributions. However, while training a conditional GAN on long-tailed data distribution, we observe that GANs suffer from model collapse on tail classes (Fig. 1). This leads to only a single pattern being generated for a particular tail class. To investigate the cause of this phenomenon, we inspect the class-specific parameters of cBN, which are gain $\gamma_{\mathbf{y}}^l$ and bias $\beta_{\mathbf{y}}^l$. In existing works, characteristics of groups of features have been insightful for analysis of neural networks and have led to development of regularization techniques [9, 40]. Hence for further analysis we also create n_g groups of the $\gamma_{\mathbf{y}}^l$ parameters and stack them to obtain $\mathbf{\Gamma}_{\mathbf{y}}^l \in \mathbb{R}^{n_g \times n_c}$, where n_c are the number of columns after grouping. It is observed that the value of spectral norm ($\sigma_{\max}(\mathbf{\Gamma}_{\mathbf{y}}^l) \in \mathbb{R}$) explodes (*i.e.* increases abnormally) as mode collapse occurs for corresponding tail class y as shown in Fig. 2. We observe this phenomenon consistently across both the smaller SNGAN [26] (Fig. 2) and the larger BigGAN [2] (Fig. 1) model. We observe similar spectral explosion for BigGAN model as in Fig. 2 (empirically shown in Fig. 9). In the earlier works, mode collapse could be detected by anomalous behaviour of spectral norm of discriminator (refer to suppl. material for details). However in the class-specific mode collapse the discriminator’s spectral norms show normal behavior and are unable to signal such collapse. Here, our analysis of $\sigma_{\max}(\mathbf{\Gamma}_{\mathbf{y}}^l)$ helps in detecting class-specific mode collapse.

3.3 Group Spectral Regularizer (gSR)

Our aim now is to prevent the class-specific mode collapse while training. To achieve this, we introduce a regularizer for the generator G which prevents spectral explosion. We would like to emphasize that earlier works including Augmentations [45], LeCam [36] regularizer etc. are applied on discriminator, hence our regularizer’s focus on G is complementary to those of existing techniques. As we observe that spectral norm explodes for the $\gamma_{\mathbf{y}}^l$ and $\beta_{\mathbf{y}}^l$, we deploy a group Spectral Regularizer (gSR) to prevent mode collapse. Steps followed by gSR for estimation of σ_{\max} of $\gamma_{\mathbf{y}}^l (\in \mathbb{R}^d)$ are described below (also given in Fig. 3):

$$\text{Grouping} : \mathbf{\Gamma}_{\mathbf{y}}^l = \Pi(\gamma_{\mathbf{y}}^l, n_g) \in \mathbb{R}^{n_g \times n_c} \quad (3)$$

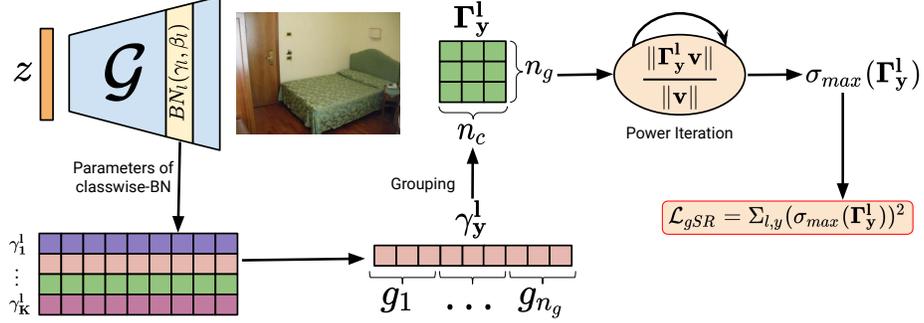


Fig. 3: Algorithmic overview. During each training step, 1) we extract the gain γ_y^1 for each cBN layers in G , 2) group them into matrix Γ_y^1 and estimate $\sigma_{\max}(\Gamma_y^1)$. 3) We repeat the same procedure with bias β_y^1 to obtain $\sigma_{\max}(\mathbf{B}_y^1)$. 4) Finally, we regularize both as described in \mathcal{L}_{gSR} (Eq. 5).

$$\text{Power Iteration : } \sigma_{\max}(\Gamma_y^1) = \max_{\mathbf{v}} \|\Gamma_y^1 \mathbf{v}\| / \|\mathbf{v}\| \quad (4)$$

$\mathbf{v} (\in \mathbb{R}^d)$ is a random vector for power iterations. n_g and n_c are the number of groups and number of columns respectively, such that $n_g \times n_c = d$. After estimation of $\sigma_{\max}(\Gamma_y^1)$ and similarly $\sigma_{\max}(\mathbf{B}_y^1)$, the regularized loss objective for generator can be written as:

$$\min_G \mathcal{L}_G + \lambda_{gSR} \mathcal{L}_{gSR}; \text{ where } \mathcal{L}_{gSR} = \sum_l \sum_y \lambda_y (\sigma_{\max}^2(\Gamma_y^1) + \sigma_{\max}^2(\mathbf{B}_y^1)) \quad (5)$$

As the spectral explosion is prominent for the tail classes, we weigh the spectral regularizer term with λ_y which has an inverse relation with number of samples n_y in class y . Prior work [3] shows that directly using $1/n_y$ can be over-aggressive hence, we use the effective number of samples (a soft version of inverse relation) formally given as (where $\alpha = 0.99$): $\lambda_y = (1 - \alpha) / (1 - \alpha^{n_y})$.

The regularized objective is used to update weights using backpropagation. Spectral regularizers are used in earlier works [37, 42] but they are applied on network weights W , whereas to the best of our knowledge, ours is the first work that proposes the regularization of the batch normalization (BatchNorm) parameters. There exist other form of techniques like Spectral Normalization and Spectral Restricted Isometry Property (SRIP) [1] regularizer, which we empirically did not find to be effective for our work (comparison in Sec. 5.3).

Decorrelation Effect (Relation with Group Whitening): Group Whitening [9] is a recent work which whitens the activation map X by grouping, normalizing and whitening using Zero-phase Component Analysis (ZCA) to obtain \hat{X} . Due to whitening, the rows of \hat{X}_g get decorrelated, which can be verified by finding the similarity of covariance matrix $\frac{1}{n_c} \hat{X}_g \hat{X}_g^T$ to a diagonal matrix. The

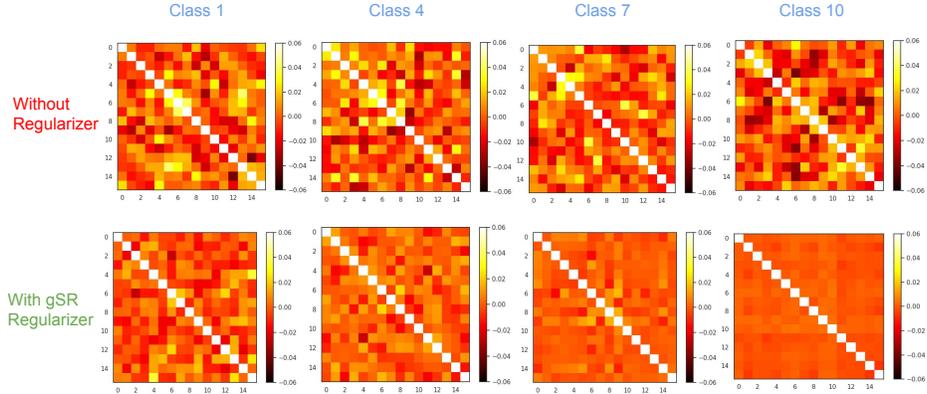


Fig. 4: Covariance matrices of Γ_y^1 for ($l = 1$) for SNGAN baseline. After using gSR (for tail classes with high λ) the covariance matrix converges to a diagonal matrix in comparison to without gSR (where large correlations exist). This demonstrates the decorrelation effect of gSR on γ_y^1 , which alleviates class-specific mode collapse.

Group Whitening transformation significantly improves the generalization performance by learning diverse features. As our regularizer also operates on groups of features, we find that minimizing the \mathcal{L}_{gSR} loss also leads to decorrelation of the rows of Γ_y^1 . We verify this phenomenon by visualizing the covariance matrix $\frac{1}{n_c}[\Gamma_y^1 - \mathbb{E}[\Gamma_y^1]]([\Gamma_y^1 - \mathbb{E}[\Gamma_y^1]])^\top$.

In Fig. 4, we plot the covariance matrices for both the SNGAN and SNGAN with regularizer (gSR). We clearly observe that for tail classes with high λ_y the covariance matrix is more closer to a diagonal matrix which confirms the decorrelation of parameters caused by gSR. We find that decorrelation is required more in layers with more class-specific information (*i.e.* earlier layers of generator) rather than layers with generic features like edges. We provide the visualizations for more layers in the suppl. material.

Recent theoretical results [11, 39] for supervised learning show that decorrelation of parameters mitigates overfitting, and leads to better generalization. This is analogous to our observation of decorrelation being able to prevent mode collapse and helpful in generating diverse images.

4 Experimental Evaluation

We perform extensive experiments on various long-tailed datasets with different resolution. For the controlled imbalance setting, we perform experiments on CIFAR-10 [18] and LSUN [43], which are commonly used in the literature [3, 4, 33] (Sec. 4.1). We also show results on challenging real-world datasets (with skewed data distribution) of iNaturalist2019 [10] and AnimalFaces [17] (Sec. 4.2).

Datasets: We use the CIFAR-10 [18] and a subset (5 classes) of LSUN dataset [43] (50k images balanced across classes) for our experiments. The choice of 5 class

Table 1: Quantitative results on the CIFAR-10 and LSUN dataset. On an average, we observe a relative improvement in FID of 20.33% and 39.08% over SNGAN and BigGAN baselines respectively.

Imb. Ratio (ρ)	CIFAR-10				LSUN			
	100		1000		100		1000	
	FID (\downarrow)	IS (\uparrow)	FID (\downarrow)	IS (\uparrow)	FID (\downarrow)	IS (\uparrow)	FID (\downarrow)	IS (\uparrow)
CBGAN [31]	33.01 \pm 0.12	6.58 \pm 0.05	44.82 \pm 0.12	5.92 \pm 0.05	37.41 \pm 0.10	2.82 \pm 0.03	44.70 \pm 0.13	2.77 \pm 0.02
LSGAN [24]	24.36 \pm 0.01	7.77 \pm 0.07	51.47 \pm 0.21	6.54 \pm 0.05	37.64 \pm 0.05	3.12 \pm 0.01	41.50 \pm 0.04	2.74 \pm 0.02
SNGAN [26]	30.62 \pm 0.07	6.80 \pm 0.02	54.58 \pm 0.19	6.19 \pm 0.01	38.17 \pm 0.02	3.02 \pm 0.01	38.36 \pm 0.11	2.99 \pm 0.01
+ gSR (Ours)	18.58 \pm 0.10	7.80 \pm 0.09	48.69 \pm 0.04	5.92 \pm 0.01	28.84 \pm 0.09	3.50 \pm 0.01	35.76 \pm 0.05	3.56 \pm 0.01
BigGAN [2]	19.55 \pm 0.12	8.80 \pm 0.09	50.78 \pm 0.23	6.50 \pm 0.05	38.65 \pm 0.05	4.02 \pm 0.01	45.89 \pm 0.30	3.25 \pm 0.01
+ gSR (Ours)	12.03 \pm 0.08	9.21 \pm 0.07	38.38 \pm 0.01	7.24 \pm 0.04	20.18 \pm 0.07	3.67 \pm 0.01	24.93 \pm 0.09	3.68 \pm 0.01

subset is for a direct comparison with related works [31, 33] which identify this subset as challenging and use that for experiments. For converting the balanced subset to the long-tailed dataset with imbalance ratio (ρ) (*i.e.* ratio of highest frequency class to the lowest frequency class), we remove the additional samples from the training set. Prior works [3, 4, 25] follow this standard process to create benchmark long-tailed datasets. We keep the validation sets balanced and unchanged to evaluate the performance by treating all classes equally. We provide additional details about datasets in the suppl. material. We perform experiments on the imbalance ratio of 100 and 1000. In case of CIFAR-10 for $\rho = 1000$ the majority class contains 5000 samples whereas the minority class has only 5 samples. For performing well in this setup, the GAN has to robustly learn from many-shot (5000 sample class) as well as at few-shot (5 sample class) together, making this benchmark challenging.

Evaluation: We report the standard Inception Score (IS) [32] and Fréchet Inception Distance (FID) metrics for the generated datasets. We report the mean and standard deviation of 3 evaluation runs similar to the protocol followed by DiffAug [45] and LeCam [36]. We use a held out set of 10k images for calculation of FID for both the datasets. The held out sets are balanced across classes for fair evaluation of each class.

Configuration: We perform experiments by modifying the PyTorch-StudioGAN implemented by Kang *et al.* [13], which serves as the baseline for our framework. We generate 32×32 sized images for the CIFAR-10 dataset and 64×64 sized images for the LSUN dataset. For the CIFAR-10 experiments, we use 5 D steps for each G step. Unless explicitly stated, we by default add the following two SOTA regularizers to obtain strong generic baselines for all the experiments (except CBGAN for which we follow exact setup described by Rangwani *et al.* [31]):

- DiffAugment [45]: We apply the differential augmentation technique with the colorization, translation, and cutout augmentations.
- LeCam [36]: LeCam regularizer prevents divergence of discriminator D by constraining its output through a regularizer term R_{LC} (ref. suppl. material).

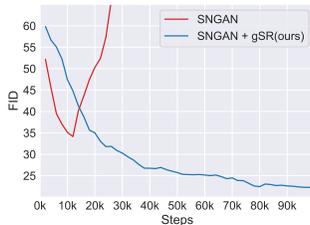


Fig. 5: Stability. Addition of gSR (to baseline) stabilizes the training to continually improve, as indicated by the FID scores.



Fig. 6: Qualitative evaluations of SNGAN baseline on LSUN dataset. Each row represents images from a class. Note the class-specific mode collapse observed in tail-classes in SNGAN (last two rows), which is alleviated after addition of gSR to generate diverse images.

Any improvement over these strong regularizers published recently is meaningful and shows the effectiveness of the proposed methods. We use a batch size of 64 for all our CIFAR-10 and LSUN experiments. For sanity check of the implementation we run the experiments for the balanced dataset (CIFAR-10) case where our FID is similar to the one obtained in LeCam [36], details are in the suppl. material.

Baselines: We compare our regularizer with the recent work of Class Balancing GAN (CBGAN) [31] which uses an auxiliary classifier for long-tailed image generation. We use the public codebase and configuration provided by the authors. The auxiliary classifiers are obtained using the LDAM-DRW as suggested by CBGAN authors. We use the SNGAN [26] (with projection discriminator [27]) as our base method on which we apply the Augmentation and LeCam regularizer for a strong baseline. We also compare our method with LSGAN [24], which is shown to be effective in preventing the mode-collapse (we use the same configuration as in SNGAN for fairness of comparison). To demonstrate improvement over large scale GANs we also use BigGAN [2] with LeCam and DiffAug regularizers as baseline. We then add our group Spectral Regularizer (gSR) in the loss terms for BigGAN and SNGAN, and report the performance in Table 1. We do not use ACGAN as our baseline as it leads to image generation which doesn't match the conditioned class label (*i.e.* class confusion) [31].

4.1 Results on CIFAR-10 and LSUN

Stability: Fig. 5 shows the FID vs iteration steps for the SNGAN and SNGAN +gSR configuration. Using gSR regularizer with SNGAN is able to effectively prevent the class-specific mode collapse, which in turn helps the GAN to improve for a long duration of time. SNGAN without gSR starts degrading quickly and stops improving very soon, this shows the stabilizing effect imparted by gSR regularizer in the training process. The stabilizing effect is similarly observed even for the BigGAN (ref. FID plot in Fig. 1).

Comparison of Quality: We observe that application of regularizer effectively avoids mode collapse and leads to a large average improvement (of 7.46) in FID for the (SNGAN + gSR) case, in comparison to SNGAN baseline across the four datasets (Table 1). Our method is also effective on BigGAN where it is able to achieve SOTA FID and IS significant improvement in almost all cases. Although SNGAN and BigGAN baselines are already enriched with SOTA regularizers of (LeCam + DiffAug) to improve results, yet the addition of our gSR regularizer significantly boosts performance by harmonizing with other regularizers. It also shows that our regularizer complements the existing work and effectively reduces mode collapse. Fig. 6 shows a comparison of the generated images for the different methods, where gSR is able to produce better images for the tail classes for LSUN dataset (refer Fig. 1 for qualitative comparison on CIFAR-10 ($\rho = 100$)). To quantify improvement over each class, we compute class-wise FID and mean FID (*i.e.* Intra FID) in Fig. 7. We find that gSR leads to very significant improvement in tail class FID as it prevents the collapse. Due to the stabilizing effect of gSR we find that head class FID are also improved, clearly demonstrating the benefit of gSR for all classes. We also provide additional metrics (precision [19], recall [19], density [29], coverage [29] and Intra-FID) in suppl. material. We find that almost all metrics show similar improvement as seen in FID (Table 1).

4.2 Results on Naturally Occurring Distributions

To show the effectiveness of our regularizer on natural distributions we perform experiments on two challenging datasets: iNaturalist-2019 [10] and AnimalFace [34]. The iNaturalist dataset is a real-world long-tailed dataset with 1010 classes of different species. There is high diversity among the images of each class, due to their distinct sources of origin. The dataset follows a long-tailed distribution with around 260k images. The second dataset we experiment with is the Animal Face Dataset [17] which contains 20 classes with with around 110 samples per class. We generate 64×64 images for both datasets using the BigGAN with a batch size of 256 for iNaturalist and 64 for AnimalFaces. The BigGAN baseline is augmented with LeCam and DiffAug regularizers. We compare our method with the baselines described in [31]. We evaluate each model using the FID on a training subset which is balanced across classes. For baselines we directly report results from Rangwani *et al.* [31] (indicated by *) in Table 2.

The BigGAN baseline achieves an FID of 6.87 on iNaturalist 2019 dataset, which improves relatively by 7.42% (-0.51 FID) when proposed gSR is combined with BigGAN. Our approach is also able to achieve FID better than SOTA CBGAN on iNaturalist 2019 dataset. Table 2 shows the performance of the BigGAN baseline over the AnimalFace dataset, where after combining with our gSR regularizer we see FID improvement by 6.90% (-2.65 FID). The improvements on both the large long-tailed dataset and few shot dataset of AnimalFace shows that gSR is able to effectively improve performance on real-world datasets. We provide additional experimental details and results in the suppl. material.

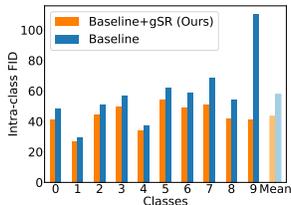


Fig. 7: Class-Wise FID and mean FID (Intra-FID) of BigGAN on CIFAR-10 over 5K generated images ($\rho = 100$).

Table 2: Quantitative results on iNaturalist-2019 and AnimalFace Dataset. We compare mean FID (\downarrow) with other existing approaches.

Method	iNaturalist 2019		AnimalFace
	cGAN	FID(\downarrow)	FID(\downarrow)
SNResGAN* [26]	✗	13.03 \pm 0.07	-
CBGAN* [31]	✗	9.01 \pm 0.08	-
ACGAN* [30]	✓	47.15 \pm 0.11	-
SNGAN* [27]	✓	21.53 \pm 0.03	-
BigGAN [2]	✓	6.87 \pm 0.04	38.41 \pm 0.04
+ gSR (Ours)	✓	6.36\pm0.04	35.76\pm0.04

5 Analysis

5.1 Ablation over Regularizers

We use the combination of existing regularizers (LeCam + DiffAug) with our regularizer (gSR) to obtain the best performing models. For further analysis of importance of each, we study their effect in comparison to gSR in this section. We perform experiments by independently applying each of them on vanilla SNGAN. Table 3 shows that existing regularizer in itself is not able to effectively reduce FID, whereas gSR is effectively able to reduce FID independently by 3.8 points. However, we find that existing regularizers along with proposed gSR, make an effective combination which further reduces FID significantly (by 9.27) on long-tailed CIFAR-10 ($\rho = 100$). This clearly shows that our regularizer effectively complements the existing regularizers.

5.2 High Resolution Image Generation

As the LSUN dataset is composed of high resolution scenes we also investigate if the class-specific mode collapse phenomenon when GANs are trained for high resolution image synthesis. For this we train SNGAN and BigGAN baselines for (128×128) using the DiffAugment and LeCAM regularizer (details in suppl. material). We find that similar phenomenon of spectral explosion leading to class-specific collapse occurs (as in 64×64 case), which is mitigated when the proposed gSR regularizer is combined with the baselines (Fig. 8). The gSR regularizer leads to significant improvement in FID (Table 4) also seen in qualitatively in Fig. 8.

5.3 Comparison with related Spectral Regularization and Normalization Techniques

As gSR constrains the exploding spectral norms for the cBN parameters (Fig. 9) to evaluate its effectiveness, we test it against other variants of spectral normalization and regularization techniques on SGAN for CIFAR-10 ($\rho = 100$).

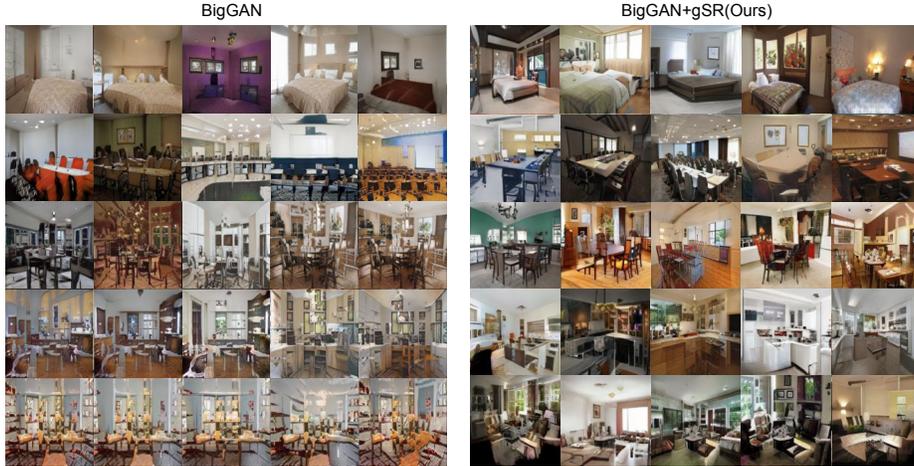


Fig. 8: Qualitative comparison of BigGAN variants on LSUN dataset ($\rho=100$) (128×128). Each row represents images from a distinct class.

Table 3: Ablation over regularizers on SNGAN. We report FID and IS on CIFAR-10 dataset (with $\rho = 100$).

LeCam+ DiffAug	gSR	FID(\downarrow)	IS(\uparrow)
\times	\times	31.73 ± 0.08	7.18 ± 0.02
\times	\checkmark	27.85 ± 0.05	7.09 ± 0.02
\checkmark	\times	30.62 ± 0.07	6.80 ± 0.02
\checkmark	\checkmark	18.58 ± 0.10	7.80 ± 0.09

Table 4: Image Generation (128×128). We report FID on both SNGAN and BigGAN on LSUN dataset (for $\rho = 100$ and $\rho = 1000$).

Imb. Ratio (ρ)	100	1000
	FID (\downarrow)	FID (\downarrow)
SNGAN [26]	53.91 ± 0.02	72.37 ± 0.08
+ gSR (Ours)	25.31 ± 0.03	31.86 ± 0.03
BigGAN [2]	61.63 ± 0.11	77.17 ± 0.18
+ gSR (Ours)	16.56 ± 0.02	45.08 ± 0.10

Group Spectral Normalization (gSN) of BatchNorm Parameters: In this setting, rather than using sum of spectral norms (Eq. 5) as regularizer for the class-specific parameters of cBN in gSR, we normalize them by dividing it by group spectral norms (*i.e.* $\frac{\gamma_y^1}{\sigma_{max}(\Gamma_y^1)}$) [26].

Group Spectral Restricted Isometry Property (gSRIP) Regularization: Extending SRIP [1], the class-specific parameters of cBN which are grouped to form a matrix Γ_y^1 , the regularizer is the sum of square of spectral norms of $(\Gamma_y^1 \Gamma_y^1 - \mathbf{I})$, (instead that of Γ_y^1 in gSR (Eq. 5)). We report our findings in Table 5. It can be inferred that all three techniques, namely gSN, gSRIP, and gSR, lead to significant improvements over the baseline. This also confirms our hypothesis that reducing (or constraining) spectral norm of cBN parameters alleviates class-specific mode collapse. However, it is noteworthy that gSR gives the highest boost over the baseline by a considerable margin in terms of FID.

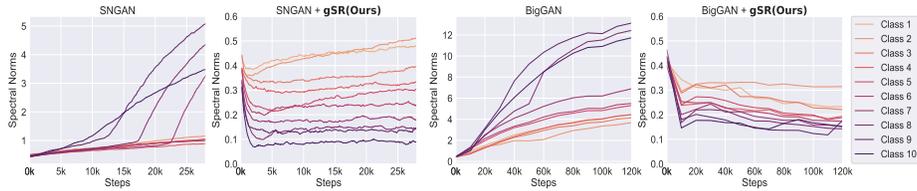


Fig. 9: Effect of gSR on spectral norms of $\Gamma_{\mathbf{y}}^1$ (CIFAR-10). We observe a spectral explosion both for SNGAN(*left*) and BigGAN(*right*) baselines of tail classes’ cBN parameters. This is prevented by addition of gSR as shown on corresponding right.

5.4 Analysis of gSR

In this section (and suppl. material) we provide ablations of gSR using long-tailed CIFAR-10 ($\rho=100$).

Can gSR work with StyleGAN-2? We train and analyze the StyleGAN2-ADA implementation available [13] on long-tailed datasets, where we find it also suffers from class-specific mode collapse.

We then implement gSR for StyleGAN2 by grouping 512 dimensional class conditional embeddings in mapping network to 16×32 and calculating their spectral norm which is added to loss (Eq. 5) as R_{gSR} . We find that gSR is able to effectively prevent the mode collapse (Fig. 10) and also results in significant improvement in FID in comparison to StyleGAN2-ADA baseline. Further analysis and results are present in suppl. material.

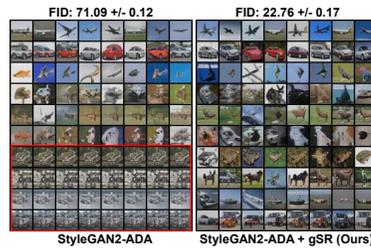


Fig. 10: StyleGAN2-ADA On CIFAR-10 ($\rho = 100$), comparison of gSR with the baseline.

What is gSR’s effect on spectral norms?

We plot spectral norms of class-specific gain parameter of 1st layer of generator in SNGAN. Spectral norms explode for the tail classes without gSR, while they remain stable when gSR is used. Fig. 5 for same experiment shows that while using gSR the FID keeps improving, whereas it collapses without using gSR. This confirms our hypothesis that constraining spectral norms stabilizes the training. We find that a similar phenomenon also occurs for BigGAN (Fig. 5) which uses SN in G , which shows that gSR is complementary to SN.

What should be the ideal number of groups? Grouping of the $(\gamma_{\mathbf{y}}^1)$ into $\Gamma_{\mathbf{y}}^1 \in \mathbb{R}^{n_g \times n_c}$ is a central operation in our regularizer formulation (Eq. 3). We group $\gamma_{\mathbf{y}}^1$ (and $\beta_{\mathbf{y}}^1$) $\in \mathbb{R}^{256}$ into a matrix $\Gamma_{\mathbf{y}}^1$ (and $\mathbf{B}_{\mathbf{y}}^1$) ablate over different combinations of n_g and n_c . Table 6 shows that FID scores do not change much significantly with n_g . As we also use power iteration to estimate the spectral norm $\sigma_{\max}(\Gamma_{\mathbf{y}}^1)$, we report iteration complexity (multiplications). Since grouping into square matrix ($n_g = 16$) gives slightly better FIDs while also being time

Table 5: Quantitative comparison of spectral regularizers. Comparison against Different Spectral Norm Regularizers on grouped cBN parameters.

	FID(↓)	IS(↑)
SNGAN [27]	30.62 \pm 0.07	6.80 \pm 0.07
+ gSN [26]	23.97 \pm 0.13	7.49 \pm 0.05
+ gSRIP [1]	23.67 \pm 0.02	7.79 \pm 0.06
+ gSR (Ours)	18.58 \pm 0.10	7.80 \pm 0.09

Table 6: Group size ablations. We report average FID and IS on CIFAR-10 dataset. $n_g = 16$ gives the best FID while also being computationally efficient, measured by per Iteration (Iter.) complexity. (Iter. complexity for power iteration method is calculated as $(n_g^2 + n_c^2)$ x number of power iterations (4 in our setting)).

n_g	n_c	FID(↓)	IS(↑)	Iter. Complexity(↓)
4	64	20.16 \pm 0.03	7.96 \pm 0.01	16448
8	32	18.69 \pm 0.06	7.80 \pm 0.01	4352
16	16	18.58 \pm 0.10	7.80 \pm 0.09	2048
32	8	20.19 \pm 0.06	7.85 \pm 0.01	4352

efficient we use it for our experiments. We also provide additional mathematical intuition for the optimality of choice of $n_c = n_g$ in suppl. material.

6 Conclusion and Future Work

In this work we identify a novel failure mode of *class-specific mode collapse*, which occurs when conditional GANs are trained on long-tailed data distribution. Through our analysis we find that the class-specific collapse for each class correlates closely with a sudden increase (explosion) in the spectral norm of its (grouped) conditional BatchNorm (cBN) parameters. To mitigate the spectral explosion we develop a novel group Spectral Regularizer (gSR), which constrains the spectral norms and alleviates mode collapse. The gSR reduces spectral norms (estimated through power iteration) of grouped parameters and leads to decorrelation of parameters, which enables GAN to effectively improve on long-tailed data distribution without collapse. Our empirical analysis shows that gSR: a) leads to improved image generation from conditional GANs (by alleviating class-specific collapse), and b) effectively complements exiting regularizers on discriminator to achieve state-of-the-art image generation performance on long-tailed datasets. One of the limitations present in our framework is that it introduces additional hyperparameter λ for the regularizer. Developing an hyperparameter free decorrelated parameterization for alleviating class-specific mode collapse is a good direction for future work. We hope that this work leads to further research on improving GANs for real-world long-tailed datasets.

Acknowledgements: This work was supported in part by SERB-STAR Project (Project:STR/2020/000128), Govt. of India and a Google Research Award. Harsh Rangwani is supported by Prime Minister’s Research Fellowship (PMRF). We thank Lavish Bansal for help with StyleGAN experiments.

References

1. Bansal, N., Chen, X., Wang, Z.: Can we gain more from orthogonality regularizations in training deep networks? *Advances in Neural Information Processing Systems* **31** (2018) 6, 12, 14
2. Brock, A., Donahue, J., Simonyan, K.: Large scale gan training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096* (2018) 1, 2, 3, 5, 8, 9, 11, 12
3. Cao, K., Wei, C., Gaidon, A., Arechiga, N., Ma, T.: Learning imbalanced datasets with label-distribution-aware margin loss. In: *Advances in Neural Information Processing Systems* (2019) 3, 6, 7, 8
4. Cui, Y., Jia, M., Lin, T.Y., Song, Y., Belongie, S.: Class-balanced loss based on effective number of samples. In: *CVPR* (2019) 3, 7, 8
5. De Vries, H., Strub, F., Mary, J., Larochelle, H., Pietquin, O., Courville, A.C.: Modulating early visual processing by language. In: *Advances in Neural Information Processing Systems* (2017) 1
6. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: *2009 IEEE conference on computer vision and pattern recognition* (2009) 3
7. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: *Advances in neural information processing systems* (2014) 1, 3
8. Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., Courville, A.C.: Improved training of wasserstein gans. In: *Advances in neural information processing systems* (2017) 3
9. Huang, L., Zhou, Y., Liu, L., Zhu, F., Shao, L.: Group whitening: Balancing learning efficiency and representational capacity. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021) 5, 6
10. iNaturalist: The inaturalist 2019 competition dataset. https://github.com/visipedia/inat_comp/tree/2019 (2019) 3, 7, 10
11. Jin, G., Yi, X., Zhang, L., Zhang, L., Schewe, S., Huang, X.: How does weight correlation affect generalisation ability of deep neural networks? *Advances in Neural Information Processing Systems* **33**, 21346–21356 (2020) 7
12. Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., Kalantidis, Y.: Decoupling representation and classifier for long-tailed recognition. In: *International Conference on Learning Representations* (2019) 3
13. Kang, M., Park, J.: Contrastive generative adversarial networks. *arXiv preprint arXiv:2006.12681* (2020) 1, 8, 13
14. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. *arXiv preprint arXiv:2006.06676* (2020) 1
15. Karras, T., Aittala, M., Hellsten, J., Laine, S., Lehtinen, J., Aila, T.: Training generative adversarial networks with limited data. In: *Proc. NeurIPS* (2020) 3
16. Kavalerov, I., Czaja, W., Chellappa, R.: cgans with multi-hinge loss. *arXiv preprint arXiv:1912.04216* (2019) 3
17. Kolouri, S., Zou, Y., Rohde, G.K.: Sliced wasserstein kernels for probability distributions. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016) 7, 10
18. Krizhevsky, A.: Learning multiple layers of features from tiny images. *Tech. rep.* (2009) 7

19. Kynkäänniemi, T., Karras, T., Laine, S., Lehtinen, J., Aila, T.: Improved precision and recall metric for assessing generative models. *Advances in Neural Information Processing Systems* **32** (2019) [10](#)
20. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017) [1](#)
21. Liu, K., Tang, W., Zhou, F., Qiu, G.: Spectral regularization for combating mode collapse in gans. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. pp. 6382–6390 (2019) [3](#)
22. Liu, M.Y., Huang, X., Yu, J., Wang, T.C., Mallya, A.: Generative adversarial networks for image and video synthesis: Algorithms and applications. *Proceedings of the IEEE* **109**(5) (2021) [3](#), [4](#)
23. Mao, Q., Lee, H.Y., Tseng, H.Y., Ma, S., Yang, M.H.: Mode seeking generative adversarial networks for diverse image synthesis. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2019) [1](#), [3](#)
24. Mao, X., Li, Q., Xie, H., Lau, R.Y., Wang, Z., Paul Smolley, S.: Least squares generative adversarial networks. In: *Proceedings of the IEEE international conference on computer vision* (2017) [8](#), [9](#)
25. Menon, A.K., Jayasumana, S., Rawat, A.S., Jain, H., Veit, A., Kumar, S.: Long-tail learning via logit adjustment. In: *International Conference on Learning Representations* (2021) [3](#), [8](#)
26. Miyato, T., Kataoka, T., Koyama, M., Yoshida, Y.: Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957* (2018) [2](#), [3](#), [5](#), [8](#), [9](#), [11](#), [12](#), [14](#)
27. Miyato, T., Koyama, M.: cGANs with projection discriminator. In: *International Conference on Learning Representations* (2018) [4](#), [5](#), [9](#), [11](#), [14](#)
28. Mullick, S.S., Datta, S., Das, S.: Generative adversarial minority oversampling. In: *The IEEE International Conference on Computer Vision (ICCV)* (2019) [3](#)
29. Naeem, M.F., Oh, S.J., Uh, Y., Choi, Y., Yoo, J.: Reliable fidelity and diversity metrics for generative models (2020) [10](#)
30. Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier gans. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70* (2017) [11](#)
31. Rangwani, H., Mopuri, K.R., Babu, R.V.: Class balancing gan with a classifier in the loop. *arXiv preprint arXiv:2106.09402* (2021) [3](#), [8](#), [9](#), [10](#), [11](#)
32. Salimans, T., Goodfellow, I., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: *Advances in neural information processing systems* (2016) [8](#)
33. Santurkar, S., Schmidt, L., Madry, A.: A classification-based study of covariate shift in gan distributions. In: *International Conference on Machine Learning*. pp. 4480–4489 (2018) [7](#), [8](#)
34. Si, Z., Zhu, S.C.: Learning hybrid image templates (hit) by information projection. *IEEE Transactions on pattern analysis and machine intelligence* **34**(7) (2011) [10](#)
35. Tran, N.T., Tran, V.H., Nguyen, N.B., Nguyen, T.K., Cheung, N.M.: On data augmentation for gan training. *IEEE Transactions on Image Processing* **30**, 1882–1897 (2021) [1](#)
36. Tseng, H.Y., Jiang, L., Liu, C., Yang, M.H., Yang, W.: Regularizing generative adversarial networks under limited data. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021) [2](#), [3](#), [5](#), [8](#), [9](#)

37. Vahdat, A., Kautz, J.: NVAE: A deep hierarchical variational autoencoder. In: Neural Information Processing Systems (NeurIPS) (2020) [2](#), [6](#)
38. Wang, Y.X., Ramanan, D., Hebert, M.: Learning to model the tail. In: Advances in Neural Information Processing Systems (2017) [3](#)
39. Wang, Z., Xiang, C., Zou, W., Xu, C.: Mma regularization: Decorrelating weights of neural networks by maximizing the minimal angles. Advances in Neural Information Processing Systems **33**, 19099–19110 (2020) [7](#)
40. Wu, Y., He, K.: Group normalization. In: Proceedings of the European conference on computer vision (ECCV) (2018) [5](#)
41. Yang, Y., Xu, Z.: Rethinking the value of labels for improving class-imbalanced learning. In: NeurIPS (2020) [3](#)
42. Yoshida, Y., Miyato, T.: Spectral norm regularization for improving the generalizability of deep learning. arXiv preprint arXiv:1705.10941 (2017) [2](#), [6](#)
43. Yu, F., Zhang, Y., Song, S., Seff, A., Xiao, J.: Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. CoRR **abs/1506.03365** (2015) [7](#)
44. Zhang, H., Zhang, Z., Odena, A., Lee, H.: Consistency regularization for generative adversarial networks. arXiv preprint arXiv:1910.12027 (2019) [3](#)
45. Zhao, S., Liu, Z., Lin, J., Zhu, J.Y., Han, S.: Differentiable augmentation for data-efficient gan training. Advances in Neural Information Processing Systems **33** (2020) [1](#), [2](#), [3](#), [5](#), [8](#)
46. Zhao, Z., Singh, S., Lee, H., Zhang, Z., Odena, A., Zhang, H.: Improved consistency regularization for gans. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 11033–11041 (2021) [3](#)
47. Zhou, B., Cui, Q., Wei, X.S., Chen, Z.M.: BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition pp. 1–8 (2020) [3](#)
48. Zhou, P., Xie, L., Ni, B., Geng, C., Tian, Q.: Omni-gan: On the secrets of cgans and beyond. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 14061–14071 (2021) [3](#)