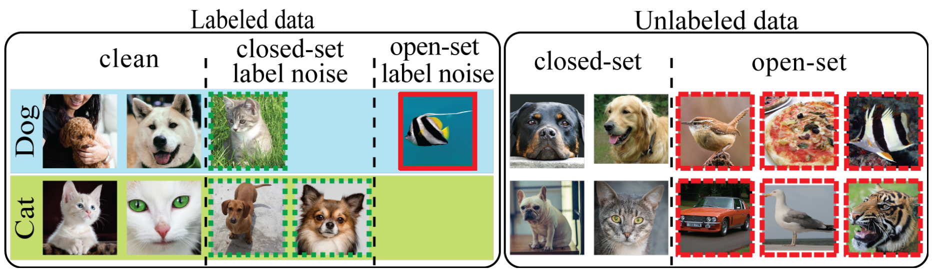
Soft Curriculum for Learning Conditional GANs with Noisy-Labeled and Uncurated Unlabeled Data

Kai Katsumata[1] Duc Minh Vo[1] Tatsuya Harada[1,2] Hideki Nakayama[1]
[1]The University of Tokyo, [2]RIKEN
{katsumata,vmduc,nakayama}@nlab.ci.i.u-tokyo.ac.jp, harada@mi.t.u-tokyo.ac.jp



	clean	closed-set label noise	open-set label noise	unlabeled closed-set	unlabeled open-set
(a) Supervised image generation	✓				
(b) Semi-supervised image generation	✓			✓	(✓)
(c) Noise robust image generation	✓	✓			
(d) Ours	✓	✓	✓	✓	✓

Figure 1. We investigate a conditional image generation in which we relax the assumption on training data. A dataset consists of labeled and unlabeled data. Labeled data contain clean samples, closed-set label noise samples whose actual categories are known classes (green dotted rectangle), and open-set label noise samples whose actual categories are outside the known classes (solid red rectangle). Unlabeled data contain closed-set samples as well as open-set samples whose categories are outside the known classes (red dashed rectangle). In contrast to previous assumptions (a,b,c), our data assumption (d) generalizes these approaches by integrating a variety of data. ✓ indicates full usage, while (✓) is partial usage.

Abstract

Label-noise or curated unlabeled data are used to compensate for the assumption of clean labeled data in training the conditional generative adversarial network; however, satisfying such an extended assumption is occasionally laborious or impractical. As a step towards generative modeling accessible to everyone, we introduce a novel conditional image generation framework that accepts noisy-labeled and uncurated unlabeled data during training: (i) closed-set and open-set label noise in labeled data and (ii) closed-set and open-set unlabeled data. To combat it, we propose soft curriculum learning, which assigns instance-wise weights for adversarial training while assigning new labels for unlabeled data and correcting wrong labels for labeled data. Unlike popular curriculum learning, which uses a threshold to pick the training samples, our soft curriculum controls the effect of each training instance by using the weights predicted by the auxiliary classifier, resulting in the preservation of useful samples while ignoring harmful ones. Our experiments show that our approach outperforms existing semi-supervised and label-noise robust methods in terms of both quantitative and qualitative performance. In particular, the proposed approach matches the performance of (semi-)supervised GANs even with less than half the labeled data.[1]

1. Introduction

Significant breakthroughs [3, 6, 17, 23, 24, 26, 34, 35, 45] in class-conditional image generation (cGANs) yield images with high fidelity and diversity; yet they are all trained in a supervised fashion where the training data consist of carefully labeled samples. However, the training data for supervised learning require immense labor cost, making it difficult to achieve a sophisticated performance. To deflate the labor cost in collecting data, semi-supervised [16, 21]

[1]The code is available at: https://github.com/raven38/NOSSGAN

and label-noise robust [14, 36] approaches have been investigated. Despite the substantial efforts of semi-supervised cGANs [16, 21] to reduce the amount of labeled data, a dataset with a high annotation cost is still required.

In this work, to significantly reduce the data collection and annotation cost, we present a new framework for training cGANs (see Fig. 1), which utilizes unreliable labeled and uncurated unlabeled data. Namely, in this study, we aim to unify the research directions for training conditional image generation on imperfect data: annotation quality [14, 36] and unannotated data [16, 21]. In our realistic data assumption, the dataset consists of two parts: noisy labeled data (*i.e.*, labeled data with closed-set and open-set label noise) and uncurated unlabeled data (*i.e.*, unlabeled data with closed-set and open-set samples). Here, closed-set and open-set label noise mean that the actual labels of samples with label noise are inside and outside the known category (label) set, respectively. Closed-set and open-set unlabeled samples also mean that the actual unknown labels are inside and outside the known category set, respectively. The objective of the new framework is to generate images with known categories. This setting generalizes (i) semi-supervised image generation [16, 21] where the labels are reliable and (ii) label-noise image generation [14, 36] where labeled data contain only closed-set label noise, and unlabeled data are not available. Hence, this new data assumption enables the use of personal collection or user-annotated data in conditional image synthesis.

To address the complex data, we propose soft curriculum learning, which makes clean and fully labeled data from noisy and partially labeled data while assigning weights to samples for adversarial training. The learning technique eliminates harmful samples (*e.g.*, samples with failed label assignment and samples far away from the training categories) while preserving useful ones (*e.g.*, samples with proper labels). Motivated by our aim, we jointly train cGAN and an auxiliary classifier that assigns clean or new labels to labeled or unlabeled samples, respectively, and confidences to all real samples. Our implicit sample selection mechanism addresses the shortcomings of curriculum learning techniques [4, 8, 43, 44], which potentially retain harmful samples and miss helpful ones because it explicitly uses a predetermined or adaptive threshold. Consequently, our approach allows to handle noisy labeled and uncurated unlabeled data naturally, resulting in maintaining the number of training samples while reducing the effects of adverse samples. Since our method is free of the hard selection procedure, we call it *soft curriculum learning*.

Our comprehensive experiments demonstrate that soft curriculum learning works well in challenging imperfect datasets containing label noise and unlabeled data. More precisely, we observe performance gains of our method over baselines in terms of the Fréchet Inception Distance (FID) [11], Inception Score (IS) [32], $F_{1/8}$, F_8 [31], and intra-FID (iFID). Qualitative results also indicate the effectiveness of our method in terms of fidelity and diversity.

In summary, our main contributions are as follows:

1. We introduce a new problem: conditional image generation trained on datasets that consist of labeled data with closed-set and open-set label noise and unlabeled data composed of closed-set and open-set samples.

2. We develop a soft curriculum technique for correcting wrong labels and assigning temporal labels while weighting the importance of each instance by employing an additional classifier trained jointly.

3. We consistently demonstrate the effectiveness of our method in experiments on various GAN architectures (*i.e.*, projection- and classifier-based cGANs) and datasets. Note that recent attempts at limited data employ only a projection GAN.

2. Related work

Conditional image generation with imperfect data. One of the prominent research directions in image generation is to build a training framework without requiring large and curated datasets [13, 15, 33, 38, 39, 48]. Semi-supervised learning approaches [5, 16, 21] explore cGANs in partially labeled data. Introducing an additional classifier enables a discriminator to be trained on labeled real data. OSSGAN [16] considers a more practical scenario where the labeled and unlabeled data do not share the label space, and it proposes entropy regularization to identify open-set samples smoothly. Robust learning for image generation [14, 36] is aimed at learning a clean conditional distribution, even when labels are noisy, by modeling a noise transition. In this study, we extend these directions to a real-world scenario. Our setting relaxes the assumption of label reliability in a semi-supervised fashion and allows robust learning to exploit open-set label noise and unlabeled data.

Semi-supervised and robust learning in image recognition. In image recognition, there also remains the issue that supervised learning requires datasets, *i.e.*, cleanly labeled large-scale datasets, which are difficult and sometimes impossible to collect. To address the issue, two popular frameworks (*i.e.*, semi-supervised [10, 27] and label-noise robust learning [25]) have been explored in recent decades. Recent attempts address a more realistic scenario where the categories of samples are not bounded by the known categories. Open-set semi-supervised learning [22, 30, 43] involves unlabeled data containing samples with categories unseen in labeled data, with the aim of classifying closed-set samples precisely while rejecting open-set samples. Learning methods robust to closed-set and open-set label noise [1, 29, 40, 42] generalize methods that only consider closed-set noise [2, 25]. In this study, we attempt to unify these research directions that are independently addressed in conditional image synthesis.

3. Problem statement

We present a novel training setting for data-efficient conditional image generation that leverages noisy labeled and uncurated unlabeled data. For K-class conditional image generation, let $\mathcal{D}_l = \{(\boldsymbol{x}_i, \boldsymbol{y}_i)\}_{i=1}^{n_l}$ be the noisy labeled training set consisting of n_l labeled samples, where a d-dimensional instance $\boldsymbol{x}_i \in \mathbb{R}^d$ and its corresponding noisy label $\boldsymbol{y}_i \in \mathcal{Y}$ sampled from labeled data distribution $p(\boldsymbol{x}, \boldsymbol{y}) : (\boldsymbol{x}_i, \boldsymbol{y}_i) \sim p(\boldsymbol{x}, \boldsymbol{y})$. The noisy label space $\mathcal{Y} = \{\boldsymbol{e}^{(1)}, \ldots, \boldsymbol{e}^{(K-1)}, \boldsymbol{e}^{(K)}\}$ consists of the standard basis vectors of the K-dimensional space. The clean label space $\bar{\mathcal{Y}} = \mathcal{Y} \cup \{\text{open-set classes}\}$ is inaccessible. Let $\mathcal{D}_u = \{\boldsymbol{u}_i\}_{i=1}^{n_u}$ be an uncurated unlabeled training set having n_u samples, where an instance $\boldsymbol{u}_i \in \mathbb{R}^d$ is sampled from the unlabeled data distribution $p(\boldsymbol{u}) : \boldsymbol{u}_i \sim p(\boldsymbol{u})$. Unlabeled data also include both closed-set and open-set samples. The goal of the conditional image generation is to model the true distribution without label noise via a generator G and a discriminator D. The generator G generates samples $G(\boldsymbol{z}, \boldsymbol{y})$ from a latent vector $\boldsymbol{z} \in \mathbb{R}^{d_z}$ and a conditioning label $\boldsymbol{y}$ drawn from a prior distribution $(\boldsymbol{z}, \boldsymbol{y}) \sim q(\boldsymbol{z}, \boldsymbol{y}) = q(\boldsymbol{z})q(\boldsymbol{y})$, where $q(\boldsymbol{z})$ is typically the standard Gaussian distribution and $q(\boldsymbol{y})$ is the uniform distribution over $\mathcal{Y}$. The discriminator D aims to identify fake samples $(G(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y})$ from real samples $(\boldsymbol{x}, \boldsymbol{y})$.

Before formulating our method, we introduce a supervised cGAN model. The conditional GANs for a fully and cleanly labeled dataset optimize the losses $\mathcal{L}_D$ and $\mathcal{L}_G$ for the discriminator and the generator, respectively:

$$\begin{aligned}\mathcal{L}_D =& \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim p(\boldsymbol{x},\boldsymbol{y})}[f_D(-D(\boldsymbol{x}, \boldsymbol{y}))] \\ &+ \mathbb{E}_{(\boldsymbol{z},\boldsymbol{y})\sim q(\boldsymbol{z},\boldsymbol{y})}[f_D(D(G(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y}))], \quad (1)\\ \mathcal{L}_G =& \mathbb{E}_{(\boldsymbol{z},\boldsymbol{y})\sim q(\boldsymbol{z},\boldsymbol{y})}[-D(G(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y})], \quad (2)\end{aligned}$$

where $f_D(\cdot) = \max(0, 1 + \cdot)$, which is a hinge loss [20, 37] for the discriminator. We alternately update the generator's and discriminator's parameters using $\mathcal{L}_G$ and $\mathcal{L}_D$ to learn a generator that generates indistinguishable samples and a discriminator that distinguishes fake and real samples well. To build our method, we customize Eqs. (1) and (2).

Although SoTA cGANs achieve outstanding performance, the absence of a dataset with sufficient quantity and reliable labels leads to poor performance and training instability. Difficulties in training on a dataset with limited quantity and quality are how to improve the stability of the training and how to estimate appropriate labels for unlabeled data under noisy labels. To overcome these difficulties, we consider a technique that assigns labels while handling label noise, based on curriculum learning and robust learning.

4. Method

Intuitive idea. Curriculum learning [4, 8, 43, 44] filters out adverse samples from the dataset, aiming to train a model on only useful samples. However, since curriculum learning employs explicit thresholds, it does not leverage the feature of ignored samples, resulting in shrinking training datasets. Furthermore, curriculum learning methods [43] for semi-supervised learning maintain label noise.

To overcome these flaws, we consider a safer way of learning cGANs on noisy data, aiming to reduce the adverse effect of misclassification while maintaining the amount of training data. Therefore, we must achieve three objectives: handling label noise containing open-set noise; handling unlabeled data including open-set samples; and eliminating samples causing negative effects from both labeled and unlabeled data. Our main idea is to make clean data from noisy labeled and uncurated unlabeled data and to control the effects of each instance tolerantly. Our method can train the discriminator on all samples via the instance-wise weight distribution, label correction, and label assignment (Fig. 2), unlike curriculum learning, which picks unlabeled samples and trains a model on all the labeled and selected unlabeled data. Our instance-wise weighting mechanism leads to a reduction of the negative effects of label noise in labeled data by assigning small weights for samples that could not be corrected by the auxiliary classifier or are open-set.

Overall concept. In addition to a generator $G : \mathbb{R}^{d_z} \times \mathcal{Y} \to \mathbb{R}^d$ and a discriminator $D : \mathbb{R}^d \times \Delta^{K-1} \to \mathbb{R}$, we employ a classifier $C : \mathbb{R}^d \to \Delta^{K-1}$ where Δ^{K-1} is a probability simplex whose vertices are in $\mathcal{Y}$. To extend the above loss function (Eqs. (1) and (2)) into our setting, we introduce discriminator losses for noisy labeled and uncurated unlabeled data, $\mathcal{L}_{\text{adv}}^{\text{lbl}} and \mathcal{L}_{\text{adv}}^{\text{unlbl}}$, respectively, and an auxiliary classifier loss $\mathcal{L}_{\text{cls}}$. Our approach can be divided into four key components: training a robust auxiliary classifier, assigning new labels to unlabeled data, correcting labels for labeled data, and weighting loss for real data (*i.e.*, both labeled and unlabeled data). To involve noisy labeled and unlabeled data, we optimize the loss functions $\mathcal{L}_D$ and $\mathcal{L}_G$:

$$\begin{aligned}\mathcal{L}_D =& \mathcal{L}_{\text{adv}}^{\text{lbl}} + \mathcal{L}_{\text{adv}}^{\text{unlbl}} + \mathcal{L}_{\text{adv}}^{\text{fake}} + \lambda \mathcal{L}_{\text{cls}}, \quad (3)\\ \mathcal{L}_G =& \mathbb{E}_{(\boldsymbol{z},\boldsymbol{y})\sim q(\boldsymbol{z},\boldsymbol{y})}[-D(G(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y})], \quad (4)\end{aligned}$$

where λ is a balancing parameter between the adversarial loss and the classification loss. We use the discriminator loss for fake data in the same way as the supervised way:

$$\mathcal{L}_{\text{adv}}^{\text{fake}} = \mathbb{E}_{(\boldsymbol{z},\boldsymbol{y})\sim q(\boldsymbol{z},\boldsymbol{y})}[f_D(D(G(\boldsymbol{z}, \boldsymbol{y}), \boldsymbol{y}))]. \quad (5)$$

Soft curriculum is an instance-wise weighting framework for discriminator training, which aims to assign small weights to harmful or irrelevant samples (*e.g.*, wrongly labeled closed-set and open-set samples) and large weights to helpful samples (*e.g.*, correctly labeled samples).

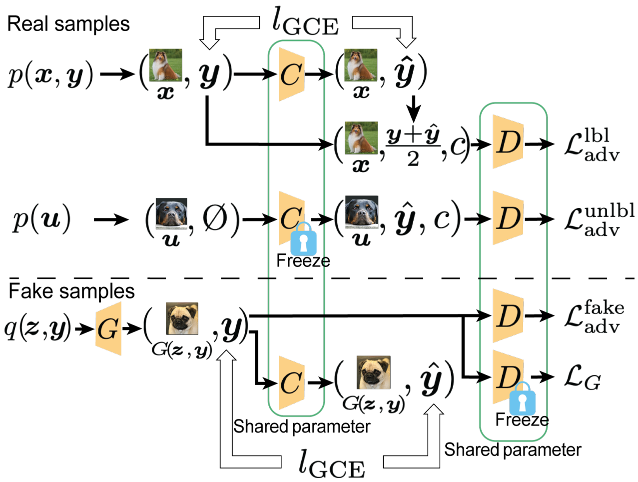


Figure 2. Overview of the proposed method. The auxiliary classifier is trained with the classification loss l_{GCE} (Eq. (7)). It corrects wrong labels in labeled samples by $C(\boldsymbol{x})$, assigns labels to unlabeled samples by $C(\boldsymbol{u})$, and distributes confidences c for discriminator optimization (Eq. (10)). The discriminator is trained with the adversarial loss for labeled, unlabeled, and fake data (Eqs. (5), (8), and (9)). Zoom in for best view.

Robust training of auxiliary classifier. We employ an auxiliary classifier for label assignment and correction (details in a later paragraph). During classifier training, besides real labeled data, we also use generated samples to increase the number of training samples. Incorporating generated samples into the training may prevent memorizing training samples (*i.e.*, overfitting). The classification loss is given by

$$\begin{aligned}\mathcal{L}_{\text{cls}} =& \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim p(\boldsymbol{x},\boldsymbol{y})}[l_{GCE}(C(\boldsymbol{x}),\boldsymbol{y})] \\ &+ \mathbb{E}_{(\boldsymbol{z},\boldsymbol{y})\sim q(\boldsymbol{z},\boldsymbol{y})}[l_{GCE}(C(G(\boldsymbol{z},\boldsymbol{y})),\boldsymbol{y})]. \end{aligned} \tag{6}$$

For robust classification with label noise, we use the generalized cross-entropy [47], which is the generalization of the mean absolute error (MAE) [7] and the cross-entropy. The loss of the generalized cross-entropy is given by

$$l_{GCE}(\hat{\boldsymbol{y}},\boldsymbol{y}) = \frac{1-(\hat{\boldsymbol{y}}^{\top}\boldsymbol{y})^{q}}{q}, \tag{7}$$

where the hyperparameter $q \in [0,1]$ controls the trade-off between optimization and noise robustness. When $q = 1$, it is equivalent to the MAE, which is robust to label noise but difficult to optimize. When $q = 0$, it is equivalent to the cross-entropy loss, which can be optimized easily. The discriminator and classifier share the feature extractor to extract features efficiently. We use the classifier prediction for the label assignment of unlabeled data and the label correction of labeled data.

Label assignment of unlabeled data. To assign new labels to unlabeled data, we take the classifier's softmax outputs $\hat{\boldsymbol{y}} = C(\boldsymbol{u})$ as a condition in discriminator inputs. We use soft labels (*i.e.*, probability vector) for the robustness to classification errors and open-set samples instead of hard labels. Soft labels prevent the discriminator inputs from wrong labels with the classifier mistake because soft labels assign a small probability to the correct class and avoid assigning a probability of 1 to the wrong class.

Label correction of labeled data. To correct noisy labels for labeled data, we take the interpolation between a given label and a predicted label, $(\boldsymbol{y}+\hat{\boldsymbol{y}})/2$, before feeding labels into the discriminator where, $\hat{\boldsymbol{y}} = C(\boldsymbol{x})$. Since some samples have proper labels depending on the label noise ratio, overwriting the given labels loses helpful information about samples with correct labels. We use the simple average because the average weighted with confidence may amplify the negative effects of wrong predictions. While we use predicted labels for inputs of the discriminator to real labeled and unlabeled samples, we maintain labels for generated samples because their labels are already proper.

Confidence assignment. To focus on helpful samples, we quantify the sample-wise importance in the discriminator training via classifier predictions. The discriminator losses for labeled and unlabeled data are defined by

$$\mathcal{L}_{\text{adv}}^{\text{lbl}} = \mathbb{E}_{(\boldsymbol{x},\boldsymbol{y})\sim p(\boldsymbol{x},\boldsymbol{y})}[cf_D(-D(\boldsymbol{x},(\boldsymbol{y}+\hat{\boldsymbol{y}})/2))], \tag{8}$$

$$\mathcal{L}_{\text{adv}}^{\text{unlbl}} = \mathbb{E}_{\boldsymbol{u}\sim p(\boldsymbol{u})}[cf_D(-D(\boldsymbol{u},\hat{\boldsymbol{y}}))], \tag{9}$$

where $\hat{\boldsymbol{y}} = C(\boldsymbol{x})$ and $\hat{\boldsymbol{y}} = C(\boldsymbol{u})$ are the softmax outputs of the classifier, and the confidence in the soft curriculum $c \in [0,1]$ is the normalized entropy of the classifier prediction:

$$c = 1 - \frac{\sum_{\hat{y}_i \in \hat{\boldsymbol{y}}} \hat{y}_i \log \hat{y}_i}{\log K}. \tag{10}$$

Here, large c is assigned for samples with high confidence and small c for samples with low confidence.

5. Experiments

Datasets. For the comprehensive evaluation, we perform experiments on TinyImageNet [41], ImageNet [28], and WebVision [19] datasets. We construct partially labeled datasets consisting of noisy labeled and uncurated unlabeled samples to benchmark our method. We use four variables that control a dataset configuration: the ratio of label noise, the number of closed-set classes, the labeled sample ratio, and the usage ratio. For the WebVision dataset, we omit the procedure for injecting label noise since it already contains label noise. To raise the open-set label noise, we first shuffle the labels by the ratio of label noise. We change a label to another label uniformly with the probability equivalent to the ratio of label noise. The label transition is run among all the classes. Second, we divide the fully labeled dataset with flipped labels partly into

Table 1. Average and standard deviation of F_8, $F_{1/8}$, FID, Inception score (IS), and iFID over three trials on TinyImageNet with 150 closed-set classes, 20% labeled samples, and 10% label noise. We compare the results of our proposed method with 15 baselines. Our method yields better performance (*i.e.*, higher F_8, $F_{1/8}$, and IS and lower FID and iFID) and consistent performance (small standard deviation). The best results are highlighted in **bold**, and the second best results are <u>underlined</u>.

	$F_8 \uparrow$	$F_{1/8} \uparrow$	FID↓	IS↑	iFID↓
DiffAug CR-GAN [48]	0.9341 ± .0103	0.9669 ± .0034	41.6848 ± 1.0075	12.0270 ± 0.3451	227.2077 ± 3.3538
RandomGAN	0.6908 ± .0310	0.8061 ± .0492	84.2262 ± 9.7936	7.6780 ± 0.6785	312.8149 ± 6.1245
SingleGAN	0.9374 ± .0009	<u>0.9761</u> ± .0018	35.5989 ± 1.5018	12.3043 ± 0.2951	233.8048 ± 4.3930
S^3GAN [21]	0.9287 ± .0027	0.9667 ± .0031	39.8652 ± 1.2017	12.1443 ± 0.2344	223.5165 ± 0.5562
OSSGAN [16]	0.8954 ± .0119	0.9598 ± .0029	46.9769 ± 3.0722	10.8745 ± 0.4495	236.6557 ± 5.0004
CurriculumGAN	0.9146 ± .0128	0.9388 ± .0144	34.4142 ± 0.6545	<u>13.3153</u> ± 0.6545	<u>217.9899</u> ± 1.5723
reRandomGAN	0.4890 ± .0396	0.7653 ± .0154	88.9622 ± 3.7217	6.8242 ± 0.5130	317.4159 ± 2.2235
reSingleGAN	0.8969 ± .0047	0.9422 ± .0099	36.2851 ± 1.3121	12.4421 ± 0.3234	237.1689 ± 2.2875
reS^3GAN	0.9089 ± .0070	0.9476 ± .0024	37.4676 ± 0.7783	13.0772 ± 0.2206	221.3113 ± 0.6992
reOSSGAN	0.8745 ± .0037	0.9320 ± .0044	40.1548 ± 1.1753	12.1081 ± 0.1531	229.1839 ± 1.6075
rcDiffAugCRGAN	0.9332 ± .0044	0.9617 ± .0078	43.5950 ± 2.2703	11.8126 ± 0.4097	226.1654 ± 4.4462
rcRandomGAN	0.7466 ± .0298	0.8801 ± .0312	69.7574 ± 4.8421	7.5598 ± 0.9622	293.7392 ± 5.9841
rcSingleGAN	<u>0.9409</u> ± .0072	0.9743 ± .0026	<u>34.1262</u> ± 1.3978	12.9476 ± 0.3931	223.1789 ± 3.8244
rcS^3GAN	0.9258 ± .0072	0.9661 ± .0056	42.0012 ± 2.1783	12.0116 ± 0.3488	228.4053 ± 4.8632
rcOSSGAN	0.9281 ± .0082	0.9692 ± .0006	42.0705 ± 1.1632	12.0458 ± 0.2670	227.5382 ± 2.3760
Ours	**0.9581** ± .0063	**0.9789** ± .0003	**29.6607** ± 0.4979	**14.7235** ± 0.3509	**206.6937** ± 2.1925

closed-set classes and partly into open-set classes. The rest of the classes, which are the number of closed-set classes subtracted from 1000 classes, are considered as open-set classes. Since label noise is brought before separation into closed-set and open-set classes, the subset for closed-set classes contains both open-set and closed-set label noise. Then, we take a subset of closed-set samples in accordance with the labeled sample ratio as labeled data, and we take the remaining closed-set samples as unlabeled data. Finally, we extract unlabeled samples from open-set class samples on the basis of the usage ratio and concatenate them with unlabeled samples that originate from closed-set samples. We use the usage ratio of 100% if not otherwise specified.

Compared methods. We use CR-BigGAN [46] with DiffAugment [48] (DiffAug CR-GAN) as a base architecture, and we build all the compared methods on it. We compare the proposed method (**Ours**) with DiffAug CR-GAN [3], RandomGAN, SingleGAN, S^3GAN [21], OSSGAN [16], and CurriculumGAN. RandomGAN is a naive baseline and assigns labels to unlabeled samples by picking a label from $\boldsymbol{y} \in \mathcal{Y}$ with equal probability. SingleGAN is another simple baseline and assigns constant labels $[1/K, \dots, 1/K]^\mathsf{T}$ to all unlabeled samples without considering their content. CurriculumGAN uses curriculum learning for semi-supervised learning by following [43] instead of our soft curriculum. For further comparison, we introduce two types of extended baseline (*i.e.*, relabeling and rcGAN [14]). The extended relabeling baselines are denoted by the prefix 're' correct labels of labeled samples by using Eq. (8) and predicted labels $\hat{\boldsymbol{y}} = C(\boldsymbol{x})$ for labeled samples. The methods with the prefix 'rc' include rcGAN, which is a technique for robust learning with label noise. The details of the compared methods are given in Supplementary Material.

Implementation details. In the experiments on the Tiny ImageNet [18] datasets at 64×64 resolution, we use a minibatch size of 1024, a latent dimension of 100, and a learning rates of 1×10^{-4} and 4×10^{-4} for the generator and the discriminator, respectively. In the experiments on the ImageNet [28] and WebVision [19] datasets at 128×128 resolution, we have a minibatch size of 256, a latent dimension of 120, and learning rates of 5×10^{-5} and 2×10^{-4} for the generator and discriminator, respectively. We update a discriminator in two steps per iteration. We train the auxiliary classifier with the same learning rate as the discriminator. We select a parameter λ in the preliminary experiments with the 150-class TinyImageNet dataset and set 0.1 for all the experiments. The parameter q in generalized cross-entropy is 0.7, which is the default value in [47].

Evaluation metrics. We use IS [32], FID [11], iFID, $F_{1/8}$ score [31], and F_8 score [31]. FID is a measure of the distance between the generated and reference images in the feature space using overall data, and iFID uses per-class data, but it was not possible to separate the evaluated values into fidelity and diversity. On the contrary, $F_{1/8}$ and F_8 quantify fidelity and diversity, respectively. We sample 10K generated images for all metrics and use the evaluation set as the reference distribution for FID, iFID, $F_{1/8}$, and F_8.

Comprehensive study. We first conduct a quantitative study on the TinyImageNet dataset with 150 closed-set classes, 50 open-set classes, 20% labeled data, and 10% la-

Table 2. Ablation study on Tiny ImageNet with 150 closed-set classes, 20% labeled samples. AB1 is the method without generalized-cross entropy. AB2 is the method without curriculum learning. AB3 is the method without curriculum for labeled data.

	30% label noise					50% label noise				
	$F_8\uparrow$	$F_{1/8}\uparrow$	FID↓	IS↑	iFID↓	$F_8\uparrow$	$F_{1/8}\uparrow$	FID↓	IS↑	iFID↓
AB1	0.8874	0.9615	36.2120	12.3104	232.8597	0.8910	0.9427	35.5164	12.2659	245.5752
AB2	0.9092	0.9619	39.7125	11.6496	236.4422	0.9131	0.9671	40.9006	10.7329	253.2198
AB3	0.9145	0.9625	31.2353	13.5164	222.1403	0.8322	0.9517	35.6693	11.4738	241.6102
Ours	**0.9238**	**0.9664**	**30.5527**	**14.0052**	**221.6443**	**0.9492**	**0.9743**	**33.0788**	**12.3833**	**238.7180**

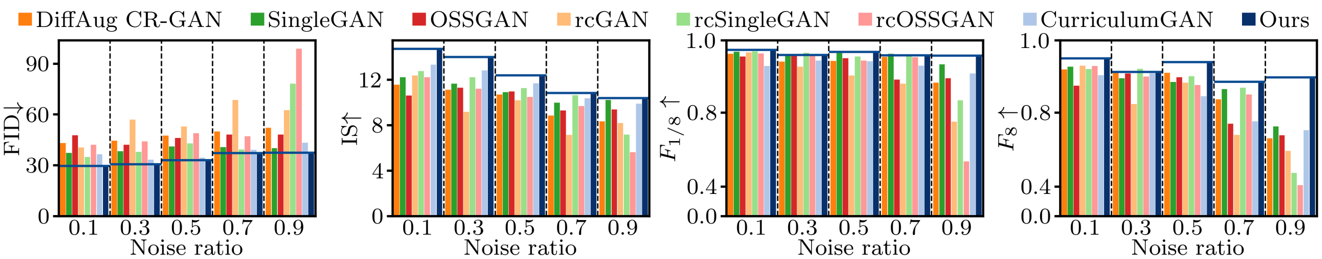


Figure 3. Quantitative results for different label noise ratios. We report the results of the experiments on the TinyImageNet dataset with 150 classes, 20% labeled data, and label noise ratios of $\{10\%, 30\%, 50\%, 70\%, 90\%\}$. We compare the methods over datasets with different label noise ratios. The blue lines indicate the results of the proposed method. Our method considerably outperforms baselines in difficult datasets (*i.e.*, high noise ratio).

Table 3. Quantitative results on ImageNet with closed-set 100 classes, 5% labeled data, 10% label noise. Our method outperforms the baselines in terms of all metrics.

	$F_8\uparrow$	$F_{1/8}\uparrow$	FID↓	IS↑	iFID↓
DiffAug CR-GAN	0.8526	0.7430	82.8757	14.6339	256.1464
RandomGAN	0.7479	0.8783	70.9336	15.0161	300.5579
SingleGAN	0.6599	0.8349	77.8994	14.0210	310.9954
S^3GAN	0.8429	0.8758	65.9445	16.1675	264.8367
OSSGAN	0.8959	0.8453	68.4343	17.4661	284.4511
Ours	**0.9443**	**0.9430**	**57.1299**	**22.3548**	**219.5597**

Table 4. Quantitative results on ImageNet with closed-set 200 classes, 5% labeled data, 10% label noise.

	$F_8\uparrow$	$F_{1/8}\uparrow$	FID↓	IS↑	iFID↓
DiffAug CR-GAN	0.8962	0.8171	56.7504	22.4951	228.9962
RandomGAN	0.7620	0.9095	49.4013	17.3114	266.7480
SingleGAN	0.7434	0.8903	51.4632	17.3922	292.1951
S^3GAN	0.4078	0.5097	111.2998	8.7617	246.3401
OSSGAN	0.9245	0.8995	44.3262	23.0263	238.2692
Ours	**0.9630**	**0.9433**	**29.6751**	**33.5418**	**183.1367**

bel noise. Namely, the dataset consists of 15K labeled samples and 85K unlabeled samples. Table 1 shows the average and standard deviation of FID, IS, $F_{1/8}$, F_8, and iFID over three trials. Our method achieves the best scores in terms of all metrics and achieves tight standard deviations, showing consistent improvement over the baselines. On the contrary, this is not the case for the improvement by rcGAN. In relabeling baselines, only classifier-based GANs improve the

Table 5. Quantitative results on ImageNet 256×256.

	$F_8\uparrow$	$F_{1/8}\uparrow$	FID↓	IS↑	iFID↓
DiffAug CR-GAN	0.8177	0.7290	83.6051	20.8947	274.5373
RandomGAN	0.7707	0.8242	60.1051	19.3663	282.4955
SingleGAN	0.8052	0.7944	62.4891	19.4504	280.3701
S^3GAN	0.9002	0.8473	52.2834	27.6553	225.6078
OSSGAN	**0.9146**	0.8124	53.7868	28.3792	229.9876
Ours	0.9076	**0.8833**	**44.5838**	**30.0695**	**214.7384**

Table 6. Quantitative results of other cGAN models on TinyImageNet. In addition to a projection-based GAN , our method shows the performance gain over classifier-based cGAN models.

	ADC-GAN [12]		TAC-GAN [9]	
	FID↓	IS↑	FID↓	IS↑
Supervised	66.5229	8.6387	50.4258	9.2594
RandomGAN	40.2519	10.5410	37.7453	10.9988
SingleGAN	43.6353	10.2666	38.5622	10.6992
S^3GAN	50.7904	10.0583	39.2887	10.5139
OSSGAN	113.1070	4.7492	41.7552	10.3462
Ours	**37.0131**	**12.1424**	**37.4393**	**11.3654**

performance from naive baselines, because reRandomGAN and reSingleGAN add much extra noise.

We then investigate the robustness of the method to label noise in experiments with different label noise ratios. In Fig. 3, we show the performance of the methods in the experiments with different label noise ratios $\{10\%, 30\%, 50\%, 70\%, 90\%\}$. Our method still outper-

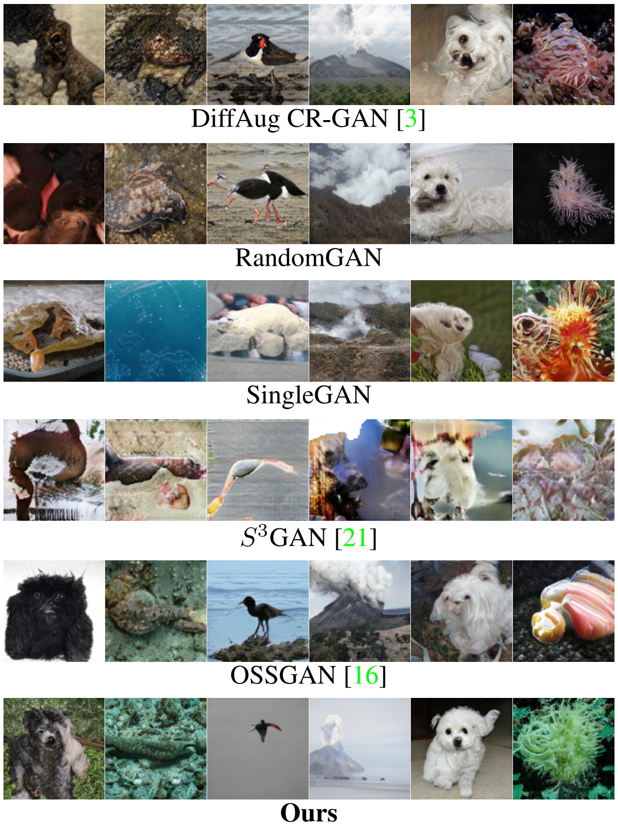


Figure 4. Class-conditional image synthesis results on ImageNet. Our method produces plausible images while respecting the given condition.

Table 7. Quantitative results on WebVision [19].

	F_8 ↑	$F_{1/8}$ ↑	FID↓	IS↑	iFID↓
DiffAug CR-GAN	0.7812	0.7725	74.3157	14.3693	249.0955
RandomGAN	0.7840	0.8627	54.8598	14.7182	246.9653
SingleGAN	0.7065	0.8276	64.8178	13.5105	280.1292
S^3GAN	0.8209	0.8680	63.4304	14.6397	238.7989
OSSGAN	0.7911	0.8294	66.7111	14.9287	242.6553
Ours	**0.8465**	**0.8866**	**51.1604**	**18.0428**	**213.5669**

forms other methods even when the labels are considerably noisy (*e.g.*, 90%). CurriculumGAN easily fails in the experiments in difficult datasets (*e.g.*, 70% or 90%).

Ablation study. To evaluate the individual contribution of each component, we carried out an ablation study. We prepare three ablation models: AB1 AB2, and AB3. AB1 is equipped with cross entropy loss instead of generalized cross-entropy, having lost the robustness to label noise. AB2 does not use curriculum learning and assigns equal weights to all samples. This method corrects wrong labels, assigns new labels, and distributes equal weights to all samples, and their classifier is trained on generalized cross-entropy. AB3 does not correct the labels of the labeled data. The method assigns new labels to unlabeled data and distributes weights in accordance with the classifier's confidence. It is close to ordinal curriculum learning. The results of the ablation study on two configurations are given in Tab. 2. With cross-entropy, AB1 drops in performance as a result of the contribution of the robust classifier. Since correcting labels of labeled samples without soft curriculum may add extra label noises, AB2 exhibits the worst performance in terms of FID, IS, and iFID in datasets with a high label noise ratio. AB3 shows a large degradation in performance under highly noisy data by maintaining label noise. In both trials, the final model (**Ours**) enhances the performance of the ablation models by the combination of robust training and soft curriculum learning.

Evaluation on large datasets. We evaluate the proposed method on more complex and challenging datasets to evaluate its stability. Table 3 shows the quantitative results of the ImageNet experiments. In the experiments, we observe the performance gains over baselines in terms of quantitative metrics. Figures 4 and 5 and Tab. 4 show the experimental results on the ImageNet dataset with 200 closed-set classes, 5% labeled data, 10% label noise, and 10% usage ratio. Namely, the dataset has about 12K labeled samples and 345K unlabeled samples. Our method outperforms all baselines with the quantitative metrics, as shown in Tab. 4. Figure 4 demonstrates the fidelity of the images generated by our method. Figure 5 shows consistency with Tab. 4 where our method generates images with high fidelity and diversity. With our soft curriculum, we observe the performance gain over baselines in difficult datasets with limited labeled samples, as shown in Fig. 6. In particular, our approach achieves a performance competitive with those of semi-supervised and supervised cGANs with 1/3 of the labeled data in terms of FID and IS (5% vs. 15%) and half of the labeled data in terms of $F_{1/8}$ and F_8 (5% vs. 10%).

To demonstrate the effectiveness of our method in high resolution datasets, we conduct experiments on ImageNet 256×256 with 200 closed-set classes, 4% labeled samples, 10% label noise, and 10% usage ratio. Table 5 shows that the proposed method outperforms the baselines consitently.

Evaluation on classifier-based cGANs. Next, we evaluate our method using different cGAN models. In the above evaluations, we build the compared method by integrating semi-supervised methods into projection-based cGANs. To evaluate the applicability of our method to other cGANs, we conduct experiments on additional base architectures of classifier-based cGANs (*i.e.*, ADC-GAN [12] and TAC-GAN [9]). Table 6 shows that our method outperforms the baselines in the ADC-GAN and TAC-GAN experiments.

Evaluation on real-world noise. Finally, we test our method on WebVision [19] to assess the effectiveness on real-world noise. WebVision is a dataset built via web queries, and so it contains real-world noise. We use 200 classes as the closed-set classes, drop 98% labels from the

Figure 5. Class-conditional image synthesis results on ImageNet. Our method constantly produces plausible images while respecting the given condition.

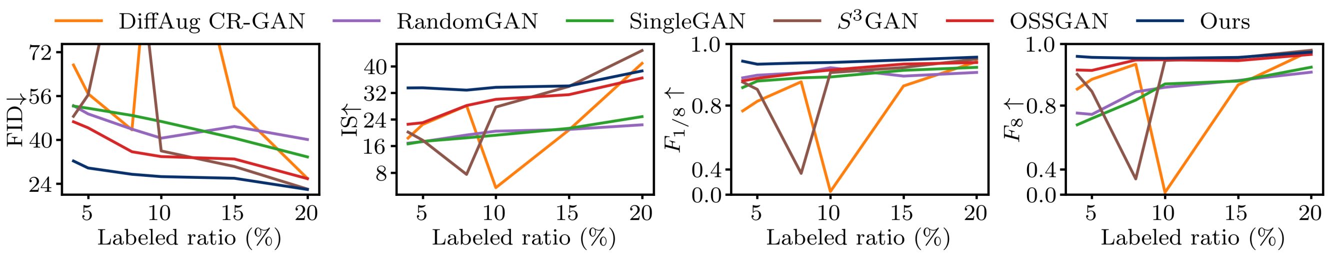


Figure 6. Quantitative results for different numbers of labeled samples. We report the results of the experiments on the ImageNet dataset with 200 classes, 10% label noise ratio, and labeled sample ratios of $\{4\%, 5\%, 8\%, 10\%, 15\%, 20\%\}$. Our method outperforms baselines in difficult datasets (blue line).

closed-set class samples to make unlabeled data, and adopt the usage ratio of 10%. Table 7 shows the results of the experiments on WebVision. We improve the performance of DiffAug CR-GAN and achieve an FID of 51.1604 with an IS of 18.0428 on the dataset with real-world noise.

6. Conclusion

We presented a novel image generation training framework that allows the training dataset to be composed of noisy labeled and uncurated unlabeled data. We proposed soft curriculum learning for this new data setting that provides clean labeled data to the discriminator while eliminating the effects of useless samples by correcting noisy labels and assigning new labels. Concurrently, we use soft labels and generalized cross-entropy loss to deal with open-set samples, avoiding overconfidence in samples that do not belong to known classes. Our comprehensive experiments show that, even when the number of labeled samples is limited and noisy, the proposed method consistently outperforms baselines in both qualitative and quantitative evaluations. Our method reduces the amount of labeled data required to achieve equivalent performance in the training of conditional GANs. Furthermore, when tested with different GAN architectures, our method demonstrates stable performance. We believe that our proposed method expands the real-world applications of cGANs in a sustainable way by making it easier to create datasets for training cGANs.

Limitation. Although our method improves baselines in challenge datasets, no beneficial improvement on datasets with sufficient labeled samples is observed. A deep analysis of the relationship between labeled data size and cGAN performance will provide further insight into the effective use of our soft curriculum method.

Acknowledgements. This work was supported by Institute for AI and Beyond of the University of Tokyo, the commissioned research (No. 225) by National Institute of Information and Communications Technology (NICT), ROIS NII Open Collaborative Research 2023_23FC01, JSPS KAKENHI Grant Numbers JP22K17947, JP23KJ0381, JP23H03449, and JP22H05015.

References

[1] Paul Albert, Diego Ortego, Eric Arazo, Noel E O'Connor, and Kevin McGuinness. Addressing out-of-distribution label noise in webly-labelled data. In *WACV*, pages 392–401, 2022. 2

[2] Dana Angluin and Philip Laird. Learning from noisy examples. *Machine Learning*, 2(4):343–370, 1988. 3

[3] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *ICLR*, 2018. 1, 5, 7

[4] Paola Cascante-Bonilla, Fuwen Tan, Yanjun Qi, and Vicente Ordonez. Curriculum labeling: Revisiting pseudo-labeling for semi-supervised learning. In *AAAI*, 2021. 2, 3

[5] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-supervised GANs via auxiliary rotation loss. In *CVPR*, pages 12146–12155, 2019. 2

[6] Kaiwen Cui, Yingchen Yu, Fangneng Zhan, Shengcai Liao, Shijian Lu, and Eric P Xing. KD-DLGAN: Data limited image generation via knowledge distillation. In *CVPR*, pages 3872–3882, 2023. 1

[7] Aritra Ghosh, Himanshu Kumar, and P Shanti Sastry. Robust loss functions under label noise for deep neural networks. In *AAAI*, 2017. 4

[8] Chen Gong, Dacheng Tao, Stephen J. Maybank, Wei Liu, Guoliang Kang, and Jie Yang. Multi-modal curriculum learning for semi-supervised image classification. *IEEE TIP*, 25(7):3249–3260, 2016. 2, 3

[9] Mingming Gong, Yanwu Xu, Chunyuan Li, Kun Zhang, and Kayhan Batmanghelich. Twin auxilary classifiers GAN, 2019. 6, 7

[10] Ryuichiro Hataya and Hideki Nakayama. Unifying semi-supervised and robust learning by mixup. In *ICLR The 2nd Learning from Limited Labeled Data (LLD) Workshop*, 2019. 2

[11] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *NeurIPS*, 2017. 2, 5

[12] Liang Hou, Qi Cao, Huawei Shen, Siyuan Pan, Xiaoshuang Li, and Xueqi Cheng. Conditional GANs with auxiliary discriminative classifier. In *ICML*, pages 8888–8902, 2022. 6, 7

[13] Liming Jiang, Bo Dai, Wayne Wu, and Chen Change Loy. Deceive D: adaptive pseudo augmentation for GAN training with limited data. pages 21655–21667, 2021. 2

[14] Takuhiro Kaneko, Yoshitaka Ushiku, and Tatsuya Harada. Label-noise robust generative adversarial networks. In *CVPR*, pages 2467–2476, 2019. 2, 5

[15] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. In *NeurIPS*, 2020. 2

[16] Kai Katsumata, Duc Minh Vo, and Hideki Nakayama. OSSGAN: open-set semi-supervised image generation. In *CVPR*, pages 11185–11193, 2022. 1, 2, 5, 7

[17] Nupur Kumari, Richard Zhang, Eli Shechtman, and Jun-Yan Zhu. Ensembling off-the-shelf models for gan training. In *CVPR*, pages 10651–10662, 2022. 1

[18] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015. 5

[19] Wen Li, Limin Wang, Wei Li, Eirikur Agustsson, and Luc Van Gool. Webvision database: Visual learning and understanding from web data. *arXiv preprint arXiv:1708.02862*, 2017. 4, 5, 7

[20] Jae Hyun Lim and Jong Chul Ye. Geometric GAN. *arXiv preprint arXiv:1705.02894*, 2017. 3

[21] Mario Lučić, Michael Tschannen, Marvin Ritter, Xiaohua Zhai, Olivier Bachem, and Sylvain Gelly. High-fidelity image generation with fewer labels. In *ICML*, pages 4183–4192, 2019. 1, 2, 5, 7

[22] Huixiang Luo, Hao Cheng, Yuting Gao, Ke Li, Mengdan Zhang, Fanxu Meng, Xiaowei Guo, Feiyue Huang, and Xing Sun. On the consistency training for open-set semi-supervised learning. *arXiv preprint arXiv:2101.08237*, 2021. 2

[23] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014. 1

[24] Takeru Miyato and Masanori Koyama. cGANs with projection discriminator. In *ICLR*, 2018. 1

[25] Nagarajan Natarajan, Inderjit S Dhillon, Pradeep K Ravikumar, and Ambuj Tewari. Learning with noisy labels. In *NeurIPS*, 2013. 2, 3

[26] Yao Ni, Piotr Koniusz, Richard Hartley, and Richard Nock. Manifold learning benefits GANs. In *CVPR*, pages 11265–11274, 2022. 1

[27] Antti Rasmus, Mathias Berglund, Mikko Honkala, Harri Valpola, and Tapani Raiko. Semi-supervised learning with ladder networks. In *NeurIPS*, 2015. 2

[28] Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, and Li Fei-Fei. ImageNet large scale visual recognition challenge. 4, 5

[29] Ragav Sachdeva, Filipe R Cordeiro, Vasileios Belagiannis, Ian Reid, and Gustavo Carneiro. EvidentialMix: learning with combined open-set and closed-set noisy labels. In *WACV*, pages 3607–3615, 2021. 2

[30] Kuniaki Saito, Donghyun Kim, and Kate Saenko. Open-Match: open-set semi-supervised learning with open-set consistency regularization. In *NeurIPS*, 2021. 2

[31] Mehdi SM Sajjadi, Olivier Bachem, Mario Lucic, Olivier Bousquet, and Sylvain Gelly. Assessing generative models via precision and recall. In *NeurIPS*, 2018. 2, 5

[32] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *NeurIPS*, 2016. 2, 5

[33] Axel Sauer, Kashyap Chitta, Jens Müller, and Andreas Geiger. Projected GANs converge faster. In *NeurIPS*, pages 17480–17492, 2021. 2

[34] Axel Sauer, Katja Schwarz, and Andreas Geiger. StyleGAN-XL: Scaling StyleGAN to large diverse datasets. In *International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2022. 1

[35] Ivan Skorokhodov, Aliaksandr Siarohin, Yinghao Xu, Jian Ren, Hsin-Ying Lee, Peter Wonka, and Sergey Tulyakov. 3D generation on ImageNet. In *ICLR*, 2022. 1

[36] Kiran K Thekumparampil, Ashish Khetan, Zinan Lin, and Sewoong Oh. Robustness of conditional GANs to noisy labels. In *NeurIPS*, 2018. 2

[37] Dustin Tran, Rajesh Ranganath, and David Blei. Hierarchical implicit models and likelihood-free variational inference. In *NeurIPS*, 2017. 3

[38] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. On data augmentation for GAN training. *IEEE TIP*, 30:1882–1897, 2021. 2

[39] Hung-Yu Tseng, Lu Jiang, Ce Liu, Ming-Hsuan Yang, and Weilong Yang. Regularizing generative adversarial networks under limited data. In *CVPR*, pages 7921–7931, 2021. 2

[40] Yisen Wang, Weiyang Liu, Xingjun Ma, James Bailey, Hongyuan Zha, Le Song, and Shu-Tao Xia. Iterative learning with open-set noisy labels. In *CVPR*, pages 8688–8696, 2018. 2

[41] Jiayu Wu, Qixiang Zhang, and Guoxi Xu. Tiny ImageNet challenge. 4

[42] Yazhou Yao, Zeren Sun, Chuanyi Zhang, Fumin Shen, Qi Wu, Jian Zhang, and Zhenmin Tang. Jo-SRC: a contrastive approach for combating noisy labels. In *CVPR*, pages 5192–5201, 2021. 2

[43] Qing Yu, Daiki Ikami, Go Irie, and Kiyoharu Aizawa. Multi-task curriculum framework for open-set semi-supervised learning. In *ECCV*, pages 438–454, 2020. 2, 3, 5

[44] Bowen Zhang, Yidong Wang, Wenxin Hou, Hao Wu, Jindong Wang, Manabu Okumura, and Takahiro Shinozaki. Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. In *NeurIPS*, 2021. 2, 3

[45] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. In *ICML*, pages 7354–7363, 2019. 1

[46] Han Zhang, Zizhao Zhang, Augustus Odena, and Honglak Lee. Consistency regularization for generative adversarial networks. In *ICLR*, 2019. 5

[47] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *NeurIPS*, 2018. 4, 5

[48] Shengyu Zhao, Zhijian Liu, Ji Lin, Jun-Yan Zhu, and Song Han. Differentiable augmentation for data-efficient GAN training. In *NeurIPS*, 2020. 2, 5