The Visual Computer
https://doi.org/10.1007/s00371-024-03360-z

RESEARCH

Image classification with consistency-regularized bad semi-supervised generative adversarial networks: a visual data analysis and synthesis

Mohammad Saber Iraji[1] · Jafar Tanha[1] · Mohammad-Ali Balafar[1] · Mohammad-Reza Feizi-Derakhshi[1]

Accepted: 5 March 2024

Abstract
Semi-supervised learning, which entails training a model with manually labeled images and pseudo-labels for unlabeled images, has garnered considerable attention for its potential to improve image classification performance. Nevertheless, incorrect decision boundaries of classifiers and wrong pseudo-labels for beneficial unlabeled images below the confidence threshold increase the generalization error in semi-supervised learning. This study proposes a novel framework for semi-supervised learning termed consistency-regularized bad generative adversarial network (CRBSGAN) through a new loss function. The proposed model comprises a discriminator, a bad generator, and a classifier that employs data augmentation and consistency regularization. Local augmentation is created to compensate for data scarcity and boost bad generators. Moreover, label consistency regularization is considered for bad fake images, real labeled images, unlabeled images, and latent space for the discriminator and bad generator. In the adversarial game between the discriminator and the bad generator, feature space is better captured under these conditions. Furthermore, local consistency regularization for good-augmented images applied to the classifier strengthens the bad generator in the generator–classifier adversarial game. The consistency-regularized bad generator produces informative fake images similar to the support vectors located near the correct classification boundary. In addition, the pseudo-label error is reduced for low-confidence unlabeled images used in training. The proposed method reduces the state-of-the-art error rate from 6.44 to 4.02 on CIFAR-10, 2.06 to 1.56 on MNIST, and 6.07 to 3.26 on SVHN using 4000, 3000, and 500 labeled training images, respectively. Furthermore, it achieves a reduction in the error rate on the CINIC-10 dataset from 19.38 to 15.32 and on the STL-10 dataset from 27 to 16.34 when utilizing 1000 and 500 labeled images per class, respectively. Experimental results and visual synthesis indicate that the CRBSGAN algorithm is more efficient than the methods proposed in previous works. The source code is available at https://github.com/ms-iraji/CRBSGAN.

Keywords Visual synthesis · Informative fake images · Low-confidence images · Bad generative adversarial network · Semi-supervised classification

1 Introduction

The capturing, recognizing, modeling, analyzing, generating, and interpreting of images have gained significant importance in the fields of artificial intelligence (AI) and machine learning [1]. They play a crucial role in understanding patterns, exploring images, and making informed decisions. These approaches employ advanced techniques specifically designed for visual data, allowing for the extraction of valuable information and the identification of complex patterns [2]. They find applications in various domains, including computer vision, medical imaging, autonomous systems, and image-based recommendation systems [3]. This is achieved by leveraging sophisticated algorithms and deep learning architectures [4].

✉ Jafar Tanha
tanha@tabrizu.ac.ir; jafar.tanha.pnu@gmail.com

Mohammad Saber Iraji
iraji.ms@gmail.com

Mohammad-Ali Balafar
Balafarila@tabrizu.ac.ir

Mohammad-Reza Feizi-Derakhshi
mfeizi@tabrizu.ac.ir

[1] Department of Computer Engineering, Faculty of Electrical and Computer Engineering, University of Tabriz, Tabriz, Iran

Published online: 06 April 2024

Supervised learning, as a subcategory of the machine learning approaches, employs significant amounts of labeled images to train predictive models for images. One of the primary limitations of model training is that manually labeling images is typically very costly and time-consuming. Creating a successful learning system is also difficult if there are only a few labeled examples available [5]. In addition, unlabeled images are typically abundant and can be obtained easily or inexpensively. Semi-supervised learning utilizes a small number of labeled images and a substantial number of unlabeled images [6], allowing the model to better capture visual features. However, the crucial question is how semi-supervised learning with unlabeled images can improve the performance of a classifier trained solely on labeled images [7].

Unlabeled images employ artificial pseudo-labels comparable to manually annotated image labels, which play a significant role in semi-supervised learning [8]. Training a model with labeled images enables self-learning methods to generate pseudo-labels for unlabeled images [9]. Among the class labels, the class with the highest predicted probability is selected as a pseudo-label [10]. Using these pseudo-labels with a probability greater than a certain threshold as reliable labels reduces the amount of error caused by incorrect pseudo-labels [11]. Nevertheless, the inability to use beneficial unlabeled images below the threshold in the learning process remains challenging [12]. The reliability of pseudo-labels for unlabeled training images and as target labels used in the consistency regularization approach is a further challenge associated with this method [13].

Consistency regularization adjusts the model to generate the same class label for input under various tolerable perturbations and augmentations [14]. Local consistency regularization involves the unification of unlabeled and labeled image labels locally and in the neighborhood of each image. Applying augmentations and perturbations to the original data is typically utilized to cover the sparse space of the data [15, 16]. The local augmentations of each image are generated by weakly augmenting the images arithmetically (such as translation and rotation for an image) [17] or by incorporating adversarial noise from the virtual adversarial training method [18]. Applying local disturbances to images near the decision boundary will result in the creation of images outside the correct boundary of the class label, and the type of local consistency regularization will decrease learning efficiency [18]. Therefore, correct consistency regularization to images with a class probability greater than the threshold (good images) and low-confidence unlabeled images contribute to improving semi-supervised classification tasks.

Recent applications of generative adversarial networks based on a generator and a discriminator network to semi-supervised learning have yielded intriguing results [19]. It is well established that using large amounts of unlabeled images for semi-supervised generative learning is essential [20]. Some adversarial learning models use the discriminator/classifier network to identify real images and predict their corresponding class labels simultaneously. Another strategy utilizes the generator and classifier networks to determine the binary distribution of the label sample [21]. In feature-matching generative adversarial networks [22], the base binary discriminator is converted into a (K + 1) class classifier to play two discriminator/classifier roles effectively. This approach has the disadvantage that the discriminator cannot effectively perform semi-supervised classification while the generator produces good fake images simultaneously. As an improvement to feature-matching generative adversarial networks, it has been reported that effective semi-supervised learning requires a "bad" generator [23]. The planned bad compliment generator could generate fake data spots in low-density regions; as a result, the classifier positioned class boundaries in these regions and augmented the generalization performance.

In contrast to the two-player game proposed in [23], the developers of marginal generative adversarial networks [24] proposed a three-player game in which the generator was encouraged to provide "bad" images for semi-supervised learning. The difficulty of marginal generative adversarial networks is that they use the maximum likelihood class prediction as the pseudo-label for all unlabeled images without label smoothing [23]. To our knowledge, consistency regularization and image augmentation for the discriminator, the bad generator, and the classifier have yet to be performed. On the other hand, the classifier using the bad generator is still incapable of detecting the correct decision boundary. Consequently, low-confidence unlabeled images with wrong pseudo-labels may negatively impact the performance of the model. Given the recent success of non-generative adversarial network-based approaches to semi-supervised learning, opportunities exist for future research to adapt semi-supervised learning elements to generative adversarial networks [25]. Solutions include consistency regularization with reliable pseudo-labeling and augmentation anchoring [26]. Based on the paragraphs above, most proposed semi-supervised classification methods lack effective utilization of learning information from unlabeled images, particularly when the probability falls below a threshold in pseudo-labeling. This issue becomes evident when the predictions of the model are uncertain or less reliable because relying solely on pseudo-labeling can result in noisy and incorrect labels. To address these challenges, we propose the incorporation of consistency regularization into a semi-supervised generative adversarial network, trained on either good or bad images, to enhance the accuracy of the semi-supervised learning model. This regularization specifically targets low-density regions near the decision boundary.

In this study, we propose a new semi-supervised classification framework that aims to enhance stability and diversity in generating bad fake images through individual consistency regularization, leading to smoother decision boundaries. Consequently, it reduces incorrect pseudo-labeling for both high-confidence and low-confidence unlabeled images, especially in cases of mislabeled images near the decision boundary. Finally, we conducted experiments to evaluate the error margin of the proposed method. The experimental results on MNIST, CIFAR-10, CINIC-10, STL-10, and SVHN datasets demonstrate that the performance of the proposed semi-supervised model is superior to that of previous research. The key contributions of our work are as follows:

1. We propose a novel framework that employs local consistency regularization to labeled, unlabeled, and latent data in three-player bad generative semi-supervised networks to improve their performance.
2. A novel type of consistency regularization loss for bad fake images, termed local consistency regularization, is introduced. The consistent bad generator efficiently learns the feature space and generates more accurate bad images (i.e., more informative images) near the true decision boundary through local consistency regularization applied to the latent space of bad fake images.
3. The local consistency regularization for good-augmented images with reliable labels applied to the classifier in the proposed framework, better adjusts the margin of the classifier for pseudo-labels generated from fake images. This action strengthens the bad generator in the generator–classifier adversarial game.
4. We demonstrate that the applied consistency regularization improves the proposed bad generative semi-supervised model, reducing the consistency-regularized semi-supervised classifier error. Reducing incorrect pseudo-labels for unlabeled images, particularly for images below the class threshold probability, lessens model error and strengthens the generalization performance of the classifier.
5. We provide a theoretical analysis of empirical risk for bad semi-supervised generative adversarial networks.
6. We demonstrate that a transformer-based discriminator provides a better signal to the bad generator in three-player bad semi-supervised generative adversarial networks.

This study is structured as follows: An introduction is provided in Sect. 1. Section 2 reviews related works. Section 3 presents the proposed semi-supervised model by generating informative fake images with consistency regularization. In Sect. 4, the experimental results of the proposed algorithm are presented. Section 5 presents a discussion, and the article ends with a conclusion.

2 Related works

This section examines previous research on semi-supervised classification and consistency regularization.

2.1 Non-generative adversarial network-based approaches to semi-supervised classification

Co-training is one of the semi-supervised algorithms that utilize pseudo-labeling [27]. The algorithm trains two classifiers for two different visions of labeled samples, and each classifier places the unlabeled samples with the highest prediction confidence into the other classifier's labeled dataset. Today, consistency regularization is widely used in the field of semi-supervised learning [28]. The teacher–student structure is the most prevalent regularization of the consistency of semi-supervised learning methods [29]. The model simultaneously learns like a student and generates labels like a teacher. The model produces potentially inaccurate targets and yields a significant error rate when applied as a learner. Reducing this risk is possible by improving the target label's quality and adjusting its generation's consistency using several techniques [30].

The ladder network [31] was the first to employ the teacher–student approach [32] that resulted from combining an encoder and a noise remover [33]. De-noising subordinates and unsupervised de-noising the error square were considered for consistency regularization in each decoder layer. Another method, the Π model, employed the propagation of the unlabeled instance forward twice in every cycle of the training process. Random data perturbation was applied to the unlabeled sample, and a random drop was input to the network layer. Forward propagations of a sample resulted in predictions that the Π model expected to have the same class [34]. Additionally, the output of the temporal ensemble idea [35] included the exponential moving average of the historical class-label predictions in different training periods. The Π model required sending samples twice per training iteration. This overhead was reduced by the temporal ensemble model's use of an exponential moving average to collect class-label predictions during the period [36].

Virtual adversarial training [37] was developed to regularize the distribution of conditional labels around any given input against local perturbations. The model must carry the same label as the original images for local perturbations surrounding each image. The local augmentation of the images near the boundary transfers the images to the other side of the class boundary; consequently, this model fails to provide the required efficiency in points near the correct class boundary. The Remixmatch method employs consistency regularization, promoting matching predictions for multiple significantly enhanced input images with those for a singular image subjected to weak augmentation [38]. Fix-match

[39] utilized "hard" labels (i.e., model output arg max) whose class probability exceeded a predefined threshold as pseudo-labels for each weak augmentation of an unlabeled instance. The model prediction was expected to be the same for this reliable weakly augmentation and strongly augmented version of the same input. Consistency regularization was not performed for pseudo-labels of unlabeled images which had confidence below the threshold limit. In addition, the challenge in the supervised part was the weak augmentation of labeled images near the decision boundary.

The authors [40] proposed DSSLDDR, a discriminative semi-supervised learning model that combines dictionary representation and deep learning to address limited labeled data. It reconstructs input data, extracts discriminative features, and balances class estimation using entropy regularization. The study also introduces DSSLDDR + , incorporating consistency/contrastive learning for improved class estimation accuracy. However, a limitation is the integration of dictionary learning only in the classification layer, limiting its potential benefits across all layers of the model.

In [41], researchers proposed dual pseudo-negative label learning (DNLL), a novel semi-supervised classification framework consisting of two sub-models that generate pseudo-negative labels for each other. This approach improves the utilization of unlabeled data and reduces parameter coupling compared to traditional methods. The study introduced a selection mechanism based on uncertainty estimation to rank the pseudo-negative labels, enhancing performance and generalization. However, addressing label quality and potential dependence on specific selection criteria are limitations of the method.

2.2 Generative adversarial network-based approaches to semi-supervised classification

Recently, semi-supervised generative learning has been evolving. A typical generative adversarial network [25] includes a generator G and a discriminator D. The objective of generator G is to learn the distribution of fake images p_g from real images p_x using noise variables with the distribution $p_z(z)$. A semi-supervised generative adversarial network simultaneously trains a generator and discriminator/classifier. Combining the loss function of an unsupervised basic generative adversarial network [42] with a supervised loss function (cross-entropy) results in the presentation of a simple semi-supervised learning method [43]. The classifier network can consist of k + 1 output units corresponding to classes $y_{1,}$ y_2, … y_{k+1}, where y_{k+1} represents the labels of the generator's images. An improved generative adversarial network solves the (K + 1) class classification problem by matching features to reduce the disparity between real and generated sample characteristics [44].

Consistency regularization for generative adversarial networks [45] is based on the improved generative adversarial network and uses a combination of local consistency, a mean teacher consistency model, and interpolation consistency [46]. Since augmentations are only performed on real images, one of the main issues with consistency regularization for generative adversarial networks is that the discriminator could "mistakenly believe" that the augmentations are real features of the target set. To circumvent this issue, regularization for generative adversarial networks [47] recommends augmenting the generated samples before they enter the discriminator so that the discriminator is uniformly regularized. The discriminator pays attention to both real and fake augmentations and thus focuses on meaningful visual information. The algorithm is implemented on the basic generative adversarial network in a two-player game with the objective of enhancing the image quality produced by good generators (high-confidence images).

Triple adversarial generative networks [48] are characterized as a three-player game. This structure has three components: a) a generator with a neural network to produce fake samples conditioned on real labels, b) a classifier that generates pseudo-labels for imported real images, and c) a discriminator that determines whether an image-label pair from the data set has a real label or not. Due to the imbalance between real and fake pairs, the discriminator tends to over-remember labeled real samples. In addition, the classifier made false predictions on unlabeled images. A class conditional generative adversarial network with random regional replacement (R3-CGAN) [21] was developed based on the triangle generative adversarial network (Triangle-GAN) [49] to address these issues. The architecture of the R3-CGAN consists of four components: 1- A generator G to generate fake images combined with given class labels, 2- A classifier C for classifying real and fake samples into k classes, 3- A discriminator (d1) to identify real or fake pairs and another discriminator (d2) to distinguish between two types of fake images. One consists of generated fake images paired with specific labels, while the other comprises an unlabeled sample paired with its pseudo-label. CutMix [24] is applied to inter-class examples and inter-real-fake samples to achieve consistency regularization. Each pair of randomly selected images is merged by replacing a rectangular region with another image. The replacement region is determined by the beta distribution of the random variable γ. Consistency regularization is based on the sample-class pairwise distribution and a good generator.

The generator and discriminator of the improved generative adversarial network had inconsistent loss functions; thus, the generator and discriminator failed to be simultaneously optimal [44]. The generator was unable to produce images that were sufficiently realistic for the semi-supervised classifier to function optimally. The authors [23] suggested that a generator was required to create fake images closely resembling real ones. Poorly generated samples necessitated the placement of the discriminator boundary between data manifolds of various categories, which reduced the discriminator's generalization error. An adversarial network with a potent bad generator would learn how to use a bad generator effectively to generate bad samples. CCS-GAN utilized unlabeled image clustering in conjunction with a bad generator to produce a more accurate discriminating boundary [50]. Choosing the appropriate distance criterion for clustering and time-consuming for high-dimensional data are the limitations of this method. Margin generative adversarial network (margin GAN) was designed to generate bad samples in a three-player structure [51]. The discriminator was trained to distinguish genuine samples from those generated by the generator. Similar to [23], the classifier attempted to increase the margin of real samples while decreasing the margin of fake ones. In contrast, the generator's objective was to provide realistic examples with large margins to deceive the classifier and discriminator simultaneously. Nevertheless, applying consistency regularization to bad GAN models can still improve semi-supervised learning in low-density areas.

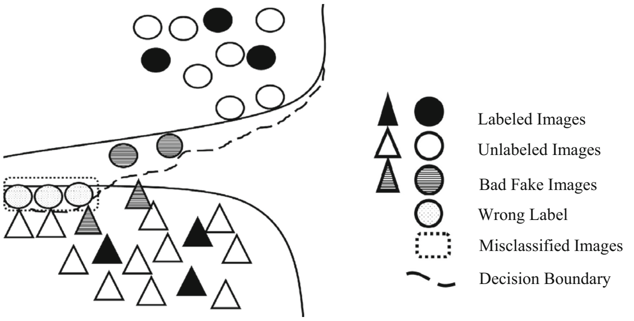


Fig. 1 Three unlabeled triangle class images were incorrectly labeled as circle pseudo-labels

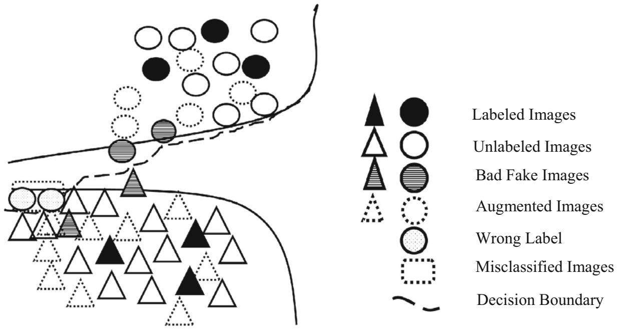


Fig. 2 Using bad images to augment images to reduce the number of erroneous pseudo-labels

3 Consistency-regularized bad semi-supervised generative adversarial networks (CRBSGAN)

The bad generator provides "informative" images near the true decision boundary with high precision, such as support vectors, and improves the generalization performance marginally [23]. In this case, in addition to the base GAN adversarial game between the discriminator and the generator [25], the generator produces images with a large margin, and the classifier aggressively makes predictions for these generated fake images with a small margin [51]. Despite efforts, the classification boundary exceeds the correct decision boundary, resulting in the mislabeling of unlabeled images. These incorrect pseudo-labels diminish the performance of the semi-supervised classifier [52]. Figure 1 demonstrates that three images have incorrect pseudo-labels and are misclassified despite border detection using a bad generator. We wish to ensure the accuracy of decision-making by combining image augmentation and consistency regularization in a bad generator (Fig. 2). The research question is, to what extent can consistency regularization with a semi-supervised generative adversarial network improve the accuracy of a semi-supervised learning model based on good or bad images?

We propose a three-layer architecture, namely consistency-regularized bad semi-supervised generative adversarial networks (CRBSGAN), consisting of a discriminator, a bad generator, and a classifier. The method incorporates weak image augmentations and consistency regularization to distinguish between fake and real images and predict pseudo-labels for unlabeled images. Local (individual) consistency regularization is applied to both the bad generator and discriminator, facilitating efficient learning of the feature space. Additionally, the model introduces self-learning-based augmentation anchoring to strengthen the classifier, particularly for good images. Figure 3 depicts the proposed model's overview. Each model component is described in detail in the sections that follow.

3.1 Regularized discriminator

In an adversarial game, the discriminator D is a deep neural network that attempts to stimulate the generator to produce images closely resembling the real distribution. Model discriminator loss functions include adversarial loss function [25] and regularization loss function [47]. In the adversarial loss function, the discriminator recognizes labeled $x^l \sim p^{\text{real}}_{x^l}$ and unlabeled images $x^u \sim p^{\text{real}}_{x^u}$ as real (labeled 1) and images produced by the generator $x^g = G(z) \sim p^{\text{fake}}_{x^g}$ as fake (labeled 0) (Eqs. 1 and 2).

We consider weak image augmentation to address data scarcity and improve the performance of deep networks [15,

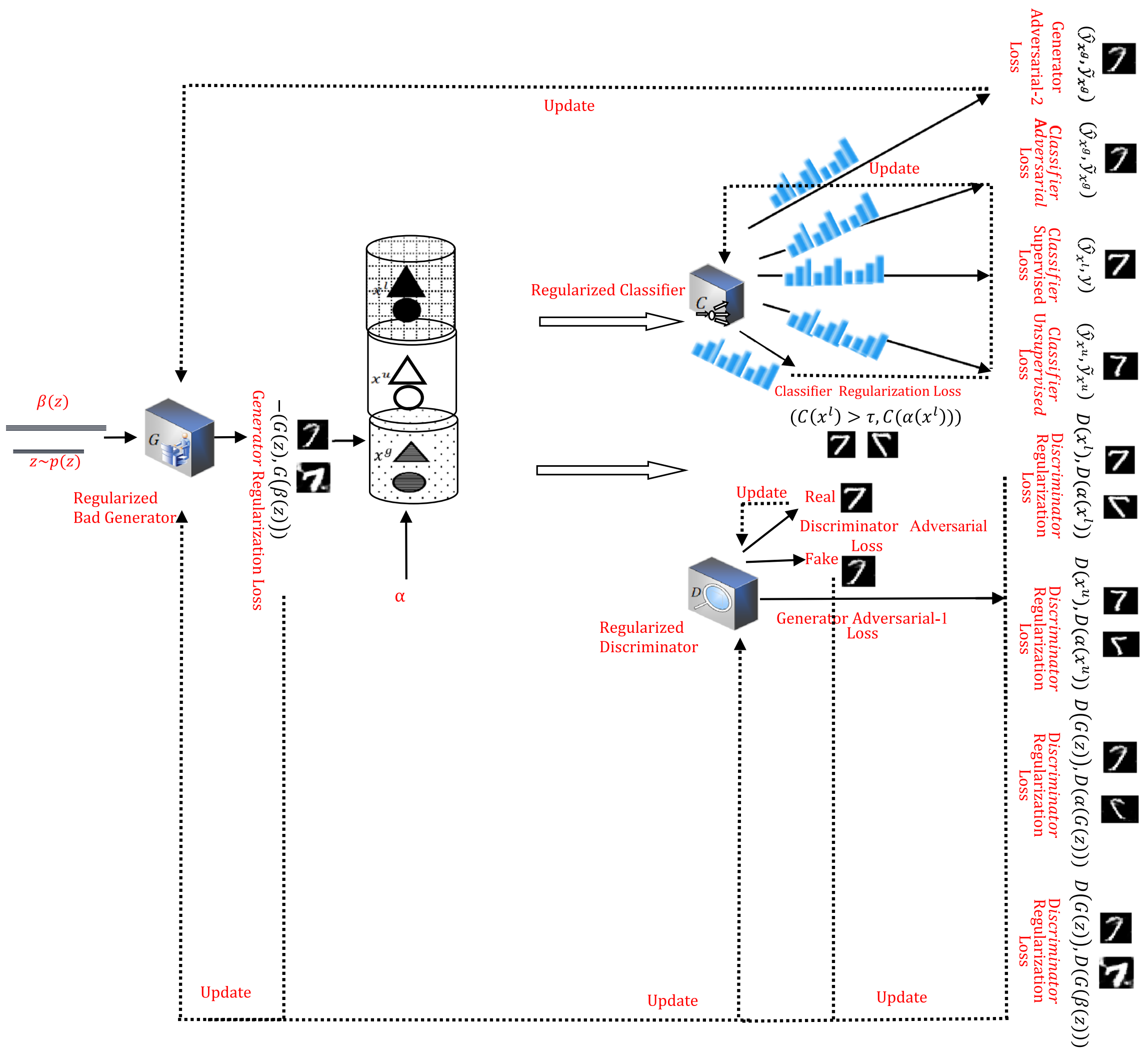


Fig. 3 Architectural overview of the CRBSGAN based on three players

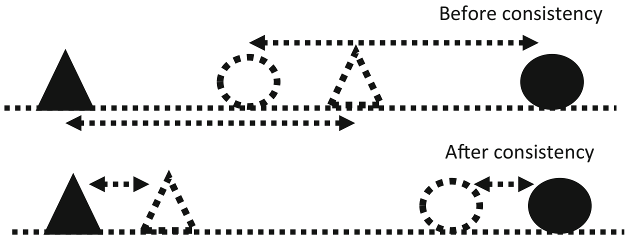


Fig. 4 The degree to which the augmented image resembles the actual image following consistent regularization in the discriminator view

16]. By applying local consistency regularization to the discriminator [53], the distance between the augmented images and their accessible original images is much closer than it was prior to consistency regularization (Fig. 4). Generated local augmentations (α is a function such as rotate) are applied to real labeled and unlabeled images, as well as fake samples. In the consistency regularization loss function, the discriminator must assign them the same label as their original images (Eq. 3). In addition, the consistency regularization loss can be computed using the $L_{2\text{norm}}$ function (Eq. 4) [54].

From the discriminator's view, images generated from the latent space and local changes to the latent space have the same label and are, therefore, fake. The β function calculates the local deviation of the sample's latent vector. β function is the addition of a vector of random numbers with a mean of μ and a variance of σ (such as the addition of a vector of random numbers with μ =0, σ =0.07 for CIFAR-10). Label consistency regularization of the aforementioned improves the learning of the feature space and the performance of the discriminator [55] and bad generator. The discriminator's final loss function is provided by Eq. 5.

$$
\begin{aligned}
L_{\text{discriminator}-\text{adversarial}} = L_{\text{CE}}\left(D\left(x^{l}\right), 1\right) \\
+ L_{\text{CE}}\left(D\left(x^{u}\right), 1\right) \\
+ L_{\text{CE}}\left(D\left(G\left(z\right)\right), 0\right)
\end{aligned} \tag{1}
$$

where

$$
\begin{aligned}
&\text{Cross Entropy (CE) Loss} = L_{\text{CE}}(y, C(x)) \\
&= -\sum_{i=1}^{nc} y_i \log\big(C(x)_i\big), \\
&\text{nc} = \text{number of class}
\end{aligned} \tag{2}
$$

$$
\begin{aligned}
&L_{\text{discriminator}-\text{consistency}-\text{regularization}} \\
&\quad = L_{\text{R}}(D(G(z)), D(G(\beta(z)))) \\
&\quad\quad + L_{\text{R}}(D(G(z)), D(\alpha(G(z)))) \\
&\quad\quad + L_{\text{R}}\Big(D\Big(x^{l}\Big), D\Big(\alpha\Big(x^{l}\Big)\Big)\Big) + L_{\text{R}}(D(x^{u}), D(\alpha(x^{u})))
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
\text{Regularization (R) Loss} &= L_{\text{R}}\ \left(y, y'\right) = L_{2\ \text{norm}}\left(y, y'\right) \\
&= \text{E}_{y\sim p_y}\ y - \text{E}_{y'\sim p_{y'}} y_2^{\prime 2}
\end{aligned} \tag{4}
$$

$$
\begin{aligned}
L_{\text{discriminator}} = {} & L_{\text{CE}}\Big(D\Big(x^{l}\Big), 1\Big) + L_{\text{CE}}\big(D\big(x^{u}\big), 1\big) + L_{\text{CE}}(D(G(z)), 0) + \lambda_1 L_{\text{R}}(D(G(z)), D(G(\beta(z)))) \\
& + \lambda_2\Big[L_{\text{R}}(D(G(z)), D(\alpha(G(z)))) + L_{\text{R}}\Big(D\Big(x^{l}\Big), D\Big(\alpha\Big(x^{l}\Big)\Big)\Big) + L_{\text{R}}\big(D\big(x^{u}\big), D\big(\alpha\big(x^{u}\big)\big)\big)\Big]
\end{aligned} \tag{5}
$$

3.2 Regularized classifier

Using consistency regularization, the multi-class classifier C is a deep neural network that attempts to predict image labels and improve the classification accuracy of semi-supervised learning. For labeled images, the classifier receives the true data label $(x^{l}, y) \sim p^{\text{real}}_{(x^{l}, y)}$ and uses the supervised loss function of cross-entropy to bring its predictions $C\big(x^{l}\big) = \widehat{y}_{x^{l}}$ closer to the true class y (Eq. 6). In fact, the margin for the labeled image class increases. The difference between the likelihood of the correct class and the maximum likelihood of the incorrect classes is referred to as the margin [56]. When the classifier makes a confident prediction, the possibility of the correct class is maximum, and the margin value is high. However, when the classifier makes an uncertain prediction, the probability distribution of the classes is flat, and the margin value is minimal [51].

$$
L_{\text{classifier}-\text{supervised}} = L_{\text{CE}}\Big(C\Big(x^{l}\Big), y\Big) \tag{6}
$$

In an adversarial game with the generator, the classifier attempts to assist in producing images with uncertain labels near the decision boundary. These images serve as a support vector in determining the predicted decision boundary. A classifier with inverse cross-entropy loss decreases the margin on the predicted label of generated images $C(G(z)) = \widehat{y}_{x^{g}}$ (Eqs. 7 and 8). For these generated examples, the pseudo-labels are considered the target label as a one-hot vector with the maximum class probability $\text{argmax}(C(G(z))) = \widetilde{y}_{x^{g}}$ [57]. Due to the limited number of labeled images, the classifier utilizes a large number of unlabeled images to improve its classification performance. The classifier with cross-entropy loss attempts to bring its predictions $C(x^{u}) = \widehat{y}_{x^{u}}$ closer to the prediction using (one-hot vector) maximum class probability $\arg\max(C(x^{u})) = \tilde{y}_{x^{u}}$ as pseudo-labels (Eq. 9).

$$
L_{\text{classifier}-\text{adversarial}} = L_{\text{ICE}}(C(G(z)), \arg\max(C(G(z)))) \tag{7}
$$

$$
\begin{aligned}
&\text{Inverse Cross Entropy (ICE) Loss} \\
&= L_{\text{ICE}}(y, c(x)) = -\sum_{i=1}^{nc} y_i \log\big(1 - c(x)_i\big), \\
&\text{nc} = \text{number of class}
\end{aligned} \tag{8}
$$

$$
L_{\text{classifier}-\text{unsupervised}} = L_{\text{CE}}\big(C\big(x^{u}\big), \arg\max\big(C\big(x^{u}\big)\big)\big) \tag{9}
$$

For reliable labeled samples, we employ consistency regularization to cover the data scarcity [58]. A reliable sample consists of data for which the predicted class probability is greater than a threshold $\mathbb{I}(\max\big(p_c\big(y|\big(x^{l}\big)\big) \geq \tau\big)$. The classifier utilizing a consistency regularization loss function endeavors to align the predicted label of a weak augmentation of the reliably labeled image with its original true label (Eq. 10). The final loss function of the classifier is found by solving Eq. 11.

$$
\begin{aligned}
&L_{\text{classifier}-\text{consistencyregularization}} \\
&= \mathbb{I}(\max\left(p_c\left(y|\left(x^{l}\right)\right) \geq \tau\right))L_{\text{R}}\left(C\left(x^{l}\right), C\left(\alpha\left(x^{l}\right)\right)\right)
\end{aligned} \tag{10}
$$

$$
L_{\text{classifier}} = L_{\text{CE}}\Big(C\Big(x^{l}\Big), y\Big) + L_{\text{ICE}}(C(G(z)), \ \arg\max(C(G(z))))
$$

$$+ L_{\text{CE}}\left(C\left(x^{u}\right), \ \arg\max\left(C\left(x^{u}\right)\right)\right)$$
$$+ \mathbb{I}(\max\left(p_{c}\left(y|\left(x^{l}\right)\right) \geq \tau\right))L_{\text{R}}(C\left(x^{l}\right), C\left(\alpha\left(x^{l}\right)\right) \tag{11}$$

3.3 Regularized bad generator

Generator G is a deep neural network with inverse convolution layers that generates bad images near the boundary. This generator engages in an adversarial game with the discriminator by attempting to make its generated images appear real to said discriminator [25]. A latent vector of z $\sim$ p_z is provided to the generator, and the parameters of the generator are updated by sending the created image (fake image $G(z)$) as the real image to the discriminator (labeled 1, whereas during training, labeled 0) (Eq. 12). The generator then attempts to produce images with a large margin in a second adversarial game with the classifier and desires the classifier to have high confidence in these images [51]. Consequently, the generator's parameters are changed so that the classification predictions on the generated images $C(G(z)) = \widehat{y}_{x^g}$ are close to their pseudo-label, i.e., the class with the highest probability $\text{argmax}(C(G(z))) = \widetilde{y}_{x^g}$, which is a one-hot vector (Eq. 13). We define the consistency regularization loss for the bad generator so that it generates distinct fake images for local latent vector deviations (Eqs. 14 and 15). Thus, the mode collapse problem [59] for the bad generator is mitigated. The combined loss terms of the bad generator are written in Eq. 16.

$$L_{\text{generator}-\text{adversarial}-1} = L_{\text{CE}}\left(\underbrace{D(\underbrace{G(z)}_{x'},}_{\hat{y}_d} \ 1\right) \tag{12}$$

$$L_{\text{generator}-\text{adversarial}-2} = L_{\text{CE}}(C(G(z)), \ \arg\max(C(G(z)))) \tag{13}$$

$$L_{\text{generator}-\text{consistency regularization}} = L_{\text{IR}}(G(z), \ G(\beta(z))) \tag{14}$$

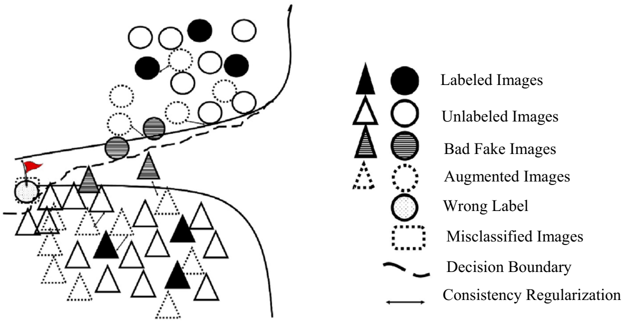


Fig. 5 Reducing the number of incorrect pseudo-labels via image augmentation and consistency regularization using bad samples

$$\begin{aligned} &\text{InverseRegularization}\,(IR)\,\text{Loss} \\ &\quad = L_{\text{IR}}\left(y, \ y^{'}\right) = -L_{\text{R}}\left(y, \ y^{'}\right) \\ &\quad = -L_{2\text{norm}}\left(y, \ y^{'}\right) - \text{E}_{y\sim p_y} y - \text{E}_{y'\sim p_{y'}} y^{'2}_{2} \end{aligned} \tag{15}$$

$$\begin{aligned} L_{\text{generator}} = {} & L_{\text{CE}}\left(\underbrace{D(\underbrace{G\,(z)}_{x'},}_{\hat{y}_d} \ 1\right) \\ & + L_{\text{CE}}\left(C\,(G\,(z))\,, \ \arg\max\,(C\,(G\,(z)))\right) \\ & + \lambda_3 L_{\text{IR}}(G\,(z)\,, \ G\,(\beta\,(z))) \end{aligned} \tag{16}$$

The consistency regularization applied to the discriminator, bad generator, and classifier facilitates the refinement of the semi-supervised model's decision boundary through the utilization of information-rich generated images (Fig. 5). Comparing Figs. 2 and 5 indicates the improved performance from the classifier view in the semi-supervised model based on the bad generator combined with image augmentation and consistency regularization. Algorithm 1 presents the pseudocode of the proposed CRBSGAN.

Algorithm 1 Consistency-regularized bad semi-supervised generative adversarial networks (CRBSGAN)

```
Inputs: $(x^l, y) \sim p^{real}_{(x^l,y)}$ :input-label pair
        $x^l \sim p^{real}_{x^l}$: real labeled images
        $x^u \sim p^{real}_{x^u}$: real unlabeled images
        z$\sim p_z$ : latent vector
        bz: batch size
        L = latent vector length
        $\theta_G$: generator parameters
        $\theta_D$: discriminator parameters
        $\theta_C$: classifier parameters
        $\alpha$ : a local augmentations function such as rotate
        $\beta$ : addition of a random number vector with a mean of $\mu$ and a variance of $\sigma$
For i = 1 of the number of training iterations, do the following:
    -Take a batch of real input-label pairs $(x^l, y) \sim p^{real}_{(x^l,y)}$, train the classifier via cross-entropy (Eq. 6), and
      update $\theta_C$.
    -Generate a batch of fake images and update the classifier parameters $\theta_C$ by pseudo labeling via inverse
      cross-entropy (Eqs. 7 and 8).
    -Take a batch of real unlabeled images $x^u \sim p^{real}_{x^u}$ , train the classifier to the maximum predicted class via
      cross-entropy (Eq. 9), and update $\theta_C$.
    -Take a hidden vector z $\sim p_z$ and generate consistency-regularized bad fake images $x^g \sim p^{fake}_{x^g}$ using the
    following steps:
            - Give a batch of the hidden vector z $\sim p_z$ to the generator and generate fake images $x^g = G(Z)$.
            - Give a batch of the generated fake images $x^g \sim p^{fake}_{x^g}$ from the step before to the discriminator,
             train it via cross-entropy (Eq. 1. term 3), which are fake images (labeled 0), and update $\theta_D$.
            - Give a batch of real images $x^u \sim p^{real}_{x^u}$ to the discriminator, train it via cross-entropy (Eq. 1. term
             2) that they are real (labeled 1), and update $\theta_D$.
            - Give a batch of real images $x^l \sim p^{real}_{x^l}$ to the discriminator, train it via cross-entropy (Eq. 1. term
             1) that they are real (labeled 1), and update $\theta_D$.
            - Give the generated fake images $x^g \sim p^{fake}_{x^g}$ of the generator to the discriminator and update the
             parameters of the generator $\theta_G$ with a real label (labeled 1) for generated fake images via cross-
             entropy (Eq. 12).
            - Update the parameters of the generator $\theta_G$ using the classifier pseudo-labeling for the fake
               images $x^g \sim p^{fake}_{x^g}$ via cross-entropy (Eq. 13).
            - Augment latent vector using $\beta$(z)
            - Update the generator parameters $\theta_G$ so that the output of the generator is inconsistent for the
             latent vector $G(z)$ and the weak augmentation of that latent vector $G(\beta(z))$ via inverse
             regularization loss (Eqs. 14 and 15).
            - Update the discriminator parameters $\theta_D$ via regularization loss (Eq. 3. term 1) so that the output
               of the discriminator is consistent for the generated images with the latent vector $D(G(z))$ and
             the generated images with weak augmentation of the latent vector $D\left(G(\beta(z))\right)$.
            - Augment local real labeled and unlabeled images using $\alpha(x^l)$ and $\alpha(x^u)$, respectively.
            - Augment local fake images via $\alpha(G(z))$
            - Update the discriminator parameters $\theta_D$ via regularization loss (Eq. 3. term 2) so that the output
               of the discriminator is consistent for the generated fake images $D(G(z))$ and the weak
             augmentation of those images $D(\alpha(G(z)))$.
            - Update the discriminator parameters $\theta_D$ via regularization loss (Eq. 3. term 3) so that the output
             of the discriminator is consistent for the real labeled images $D(x^l)$ and the weak augmentation of
             those images $D(\alpha(x^l))$.
            - Update the discriminator parameters $\theta_D$ via regularization loss (Eq. 3. term 4) so that the output
             of the discriminator is consistent for the real unlabeled images $D(x^u)$ and the weak augmentation
             of those images $D(\alpha(x^u))$.
```

4 Experiments

4.1 Data sets

To evaluate the effectiveness of the proposed semi-supervised model, we conducted experiments on three well-known datasets: MNIST [60], SVHN [61], CINIC-10 [62], and CIFAR-10 [63], STL-10 [64].

- The MNIST (Modified National Institute of Standards and Technology database) contains 60,000 training samples and 10,000 test samples consisting of handwritten digit images 0–9.
- The SVHN (Street View House Numbers) dataset is a real-world image dataset consisting of 73,257 training samples and 26,032 test samples of house numbers from 0 to 9, captured on various backgrounds.
- The CIFAR-10 (Canadian Institute for Advanced Research) dataset contains 50,000 training images and 10,000 test images, corresponding to ten classes of natural objects: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck.
- The CINIC-10 dataset expands upon CIFAR-10 by integrating images sourced from ImageNet, providing a larger-scale benchmarking option. CINIC-10 encompasses around 270,000 images and is partitioned into training, validation, and test subsets, each containing 90,000 images, making it roughly 4.5 times larger than CIFAR-10.
- The STL-10 dataset is an extension of the CIFAR-10 dataset, sharing the same ten classes as CIFAR-10. It includes 5,000 labeled training samples, 100,000 unlabeled training samples, and 8,000 test samples.

The MNIST dataset is a grayscale dataset with a single channel and images that are 28 × 28 pixels in size. On the other hand, the SVHN, CIFAR-10, CINIC-10, and STL-10 datasets consist of RGB images with three channels. The SVHN and CIFAR-10, CINIC-10 datasets have images that are 32 × 32 pixels in size, while the STL-10 dataset has higher-resolution images with a size of 96 × 96 pixels.

In semi-supervised learning, a number of the training images are labeled with their corresponding class, while the remaining training images are left unlabeled. It allows the model to learn from labeled and unlabeled images, which can improve its performance compared to using only labeled images. In the case of generative adversarial training with these datasets, a number of training images, including labels, are considered real labeled images, and the rest of the training images without labels are considered real unlabeled images. Overall, the use of these datasets in the evaluation of the proposed semi-supervised model provides a comprehensive assessment of the model's performance on a range of image classification tasks.

4.2 MNIST results

The CRBSGAN model was proposed based on the bad generator, the discriminator, and the classifier. The margin GAN [51] model, which was considered to be the base model to compare with the proposed method, also used the same networks. Figure 6 shows the architecture of the discriminator, generator, and classifier adopted from [51] for the MNIST data. The training images for 100, 600, 1000, and 3000 were labeled, while the rest were unlabeled. The learning rate for the classifier was set to 0.1, the discriminator to 0.0002, and the generator to 0.0002. The hidden vector length was 62, and the batch size was 64. The model's error rate means and deviations were evaluated on the test images over five runs. The mean error rate percentages of 2.99 ± 0.19, 2.46 ± 0.25, 2.35 ± 0.41, and 1.56 ± 0.23 were achieved using 100, 600, 1000, and 3000 labeled training images, respectively. The confusion matrices of two classifiers trained with 100 and 3000 labeled training examples on MNIST test images are depicted in Fig. 7. An accuracy of 96.54 and 98.56 was calculated for 59,900 and 57,000 unlabeled training images, respectively (Fig. 8). Additionally, Fig. 9 depicts the images created by the bad generator.

4.3 SVHN, CIFAR10 results

The proposed algorithm was executed on a computer with a 24 GB NVIDIA GeForce RTX 3090 graphics card, a 4.00 GHz Intel Core i7-6700 K processor, and 32 GB of RAM. Figure 10 depicts the discriminator and generator architecture for three-channel color images [51]. Similar to the basic article, a classifier with 12 residual blocks and Shake-Shake regularization [65] was utilized (Fig. 11). Table 1 contains the model's parameters. The dimension of the latent vector z was 100, and the batch size was set to 128. The classifier learning rate was assigned 0.05 and momentum 0.9.

We conducted experiments for 100 SVHN epochs and 150 CIFAR10 epochs. As labeled images for classifier training, 500 of the SVHN training images and 1000 and 4000 of the CIFAR10 training images were randomly selected. The error rate percentage over five runs for test images on the SVHN and CIFAR10 datasets was calculated. The mean error rate of the test images on the CIFAR10 data set with 1000 and 4000 labeled training samples, respectively, was 7.62 ± 0.35 and 4.02 ± 0.24%. The proposed method for the SVHN dataset with 500 labeled training samples achieved a mean error rate of 3.26 ± 0.11.

We conducted ablation studies on the data sets to determine the impact of the important variance hyper-parameter

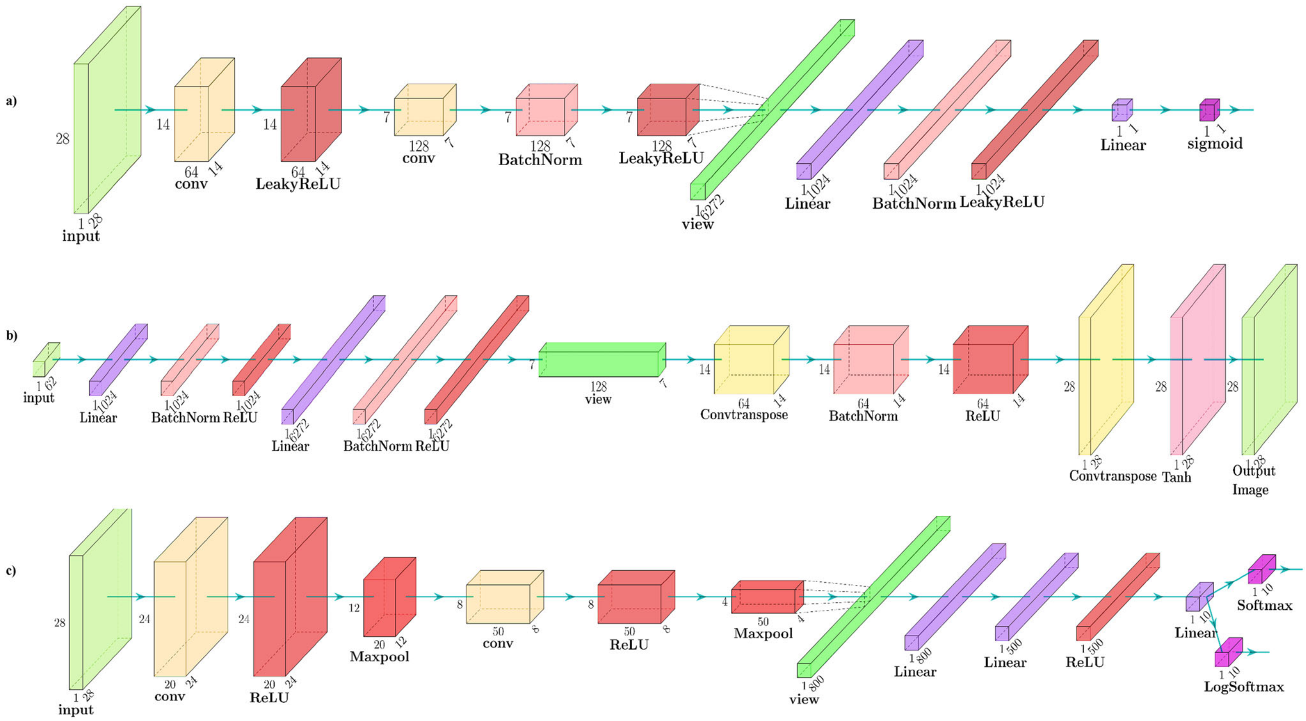


Fig. 6 The architecture of the **a** discriminator, **b** generator, and **c** classifier for the MNIST

Table 1 The proposed model parameters

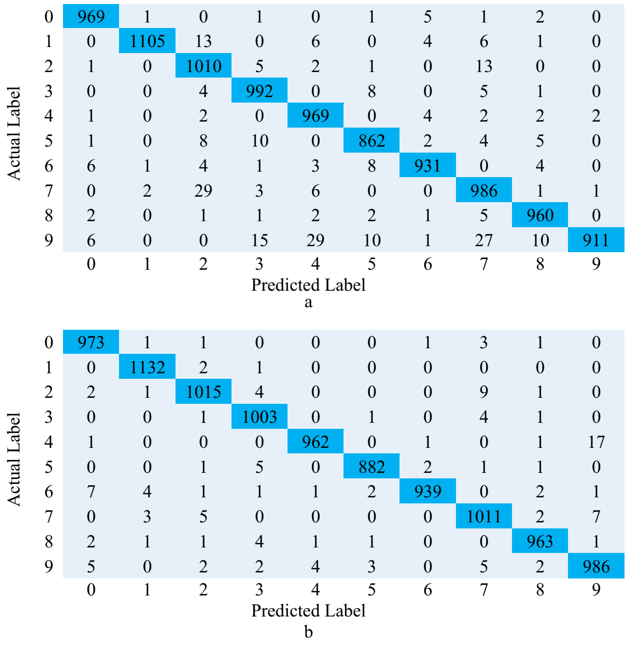
Parameters	Data				
	MNIST	SVHN	CIFAR10	CINIC-10	STL-10
lrD	0.0002	0.0002	0.0002	0.0002	0.0002
lrG	0.0002	0.0002	0.0002	0.0002	0.0002
lrC	0.1	0.05	0.05	0.1	0.1
Momentum	0.5	0.9	0.9	0.9	0.9
Beta1	0.5	0.5	0.5	0.5	0.5
Beta2	0.999	0.999	0.999	0.999	0.999
Batch_size	64	128	128	128	100
β function:(μ, σ)	(0,1)	(0,0.03)	(0,0.07)	(0,1)	(0,1)
z_dim	62	100	100	100	100
epochs	50	100	150	400	50

The best results are highlighted by bolding the values

(σ) on the latent space of a bad generator. Other neural network parameters (including layer type, training epochs, and filter size) in the proposed model were kept constant, and the error as a performance indicator was measured. The variance (σ) values versus error of the CRBSGAN on the SVHN and CIFAR10 datasets with 500 and 4000 labeled training samples, respectively, are depicted in Fig. 12. On the SVHN and CIFAR10 datasets, we observed that σ =0.03 and 0.07 are the optimal values, resulting in less classification error. Additionally, Fig. 13 depicts the fake images generated by the bad generator in conjunction with consistency regularization for the SVHN and CIFAR10 datasets.

4.4 CINIC-10 results

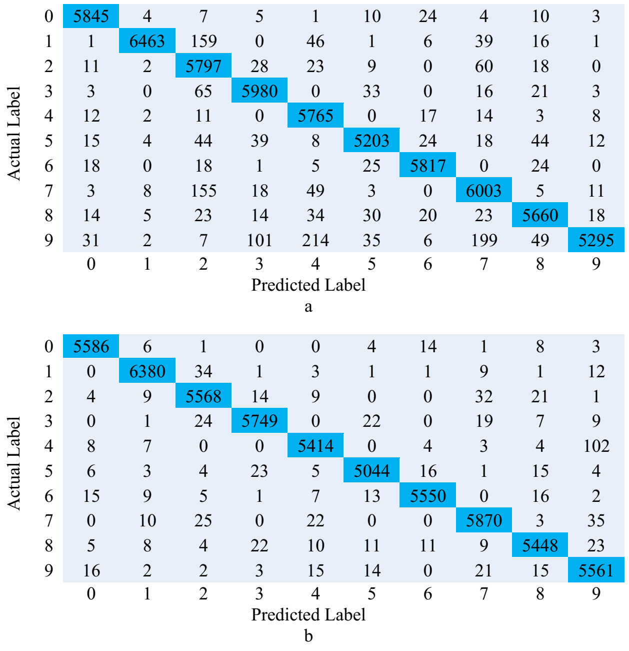
The classifier training for the proposed model on the CINIC-10 dataset involved selecting 7,000 and 10,000 labeled training images. The performance of the CNN-13 classifiers on the CINIC-10 test set is depicted in Fig. 14 through the confusion matrices. Evaluation of the classifier trained with 7,000 labeled images on the CINIC-10 test data revealed


Actual Label
Predicted Label
a
Actual Label
Predicted Label
b

	0	1	2	3	4	5	6	7	8	9
0	969	1	0	1	0	1	5	1	2	0
1	0	1105	13	0	6	0	4	6	1	0
2	1	0	1010	5	2	1	0	13	0	0
3	0	0	4	992	0	8	0	5	1	0
4	1	0	2	0	969	0	4	2	2	2
5	1	0	8	10	0	862	2	4	5	0
6	6	1	4	1	3	8	931	0	4	0
7	0	2	29	3	6	0	0	986	1	1
8	2	0	1	1	2	2	1	5	960	0
9	6	0	0	15	29	10	1	27	10	911

	0	1	2	3	4	5	6	7	8	9
0	973	1	1	0	0	0	1	3	1	0
1	0	1132	2	1	0	0	0	0	0	0
2	2	1	1015	4	0	0	0	9	1	0
3	0	0	1	1003	0	1	0	4	1	0
4	1	0	0	0	962	0	1	0	1	17
5	0	0	1	5	0	882	2	1	1	0
6	7	4	1	1	1	2	939	0	2	1
7	0	3	5	0	0	0	0	1011	2	7
8	2	1	1	4	1	1	0	0	963	1
9	5	0	2	2	4	3	0	5	2	986

Fig. 7 Confusion matrices for classifiers trained with **a** and **b** 100 and 3000 labeled training images on the MNIST test images


Actual Label
Predicted Label
a
Actual Label
Predicted Label
b

	0	1	2	3	4	5	6	7	8	9
0	5845	4	7	5	1	10	24	4	10	3
1	1	6463	159	0	46	1	6	39	16	1
2	11	2	5797	28	23	9	0	60	18	0
3	3	0	65	5980	0	33	0	16	21	3
4	12	2	11	0	5765	0	17	14	3	8
5	15	4	44	39	8	5203	24	18	44	12
6	18	0	18	1	5	25	5817	0	24	0
7	3	8	155	18	49	3	0	6003	5	11
8	14	5	23	14	34	30	20	23	5660	18
9	31	2	7	101	214	35	6	199	49	5295

	0	1	2	3	4	5	6	7	8	9
0	5586	6	1	0	0	4	14	1	8	3
1	0	6380	34	1	3	1	1	9	1	12
2	4	9	5568	14	9	0	0	32	21	1
3	0	1	24	5749	0	22	0	19	7	9
4	8	7	0	0	5414	0	4	3	4	102
5	6	3	4	23	5	5044	16	1	15	4
6	15	9	5	1	7	13	5550	0	16	2
7	0	10	25	0	22	0	0	5870	3	35
8	5	8	4	22	10	11	11	9	5448	23
9	16	2	2	3	15	14	0	21	15	5561

Fig. 8 Confusion matrices via classifiers **a** and **b** for 59,900 and 57,000 unlabeled training images on the MNIST

accurate detection values of 8092, 7479, 7345, 6980, 7132, 5982, 8236, 7887, 7853, and 7469 for the ten output classes. In contrast, the classifier trained with 10,000 labeled images exhibited enhanced detection values of 8032, 7551, 7703, 7143, 7277, 6608, 8337, 8041, 8018, and 7495 for the corresponding classes. Additionally, Fig. 15 illustrates the bad

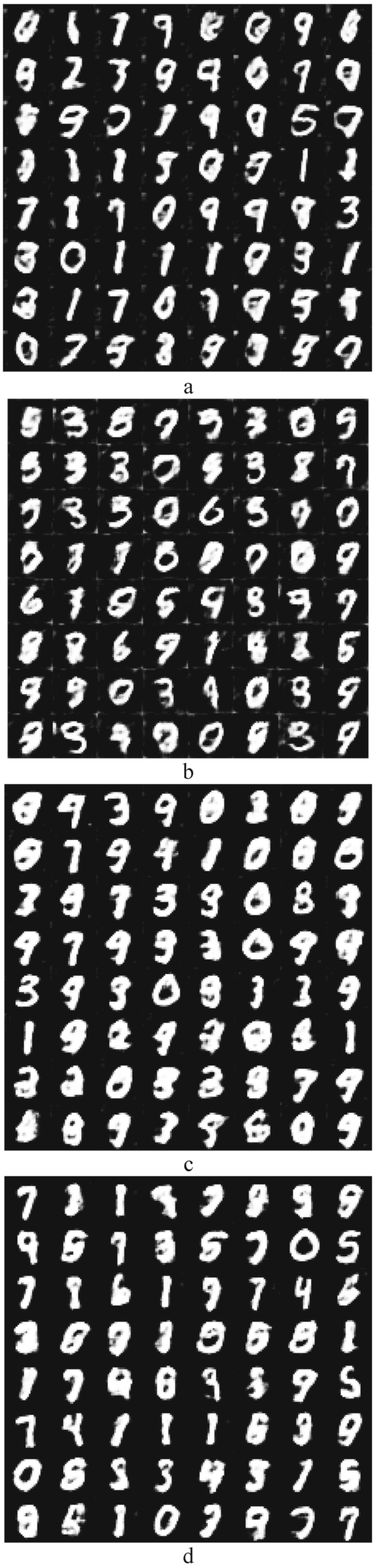


Fig. 9 Fake images generated by bad generators **a**, **b**, **c**, and **d** on the MNIST dataset where classifiers were trained with 100, 600, 1000, and 3000 labeled training images, respectively

generated images that were used for the classifiers. Furthermore, Fig. 16 displays the accuracy curve obtained from the CINIC-10 data. This curve provides visual evidence that supports the model's convergence.

4.5 STL-10 results

Previous experiments in the preceding sections were conducted on low-resolution images (32 × 32 pixels). However, it should be noted that this choice of resolution was made to demonstrate the effectiveness and efficiency of our approach in a controlled experimental setting, in line with previous studies in the field. Therefore, it does not imply that our proposed method is limited to such images. To provide a more comprehensive analysis, we performed additional experiments on higher-resolution images with a resolution of 96 × 96 pixels. These new experiments were conducted on the STL10 dataset and aimed to showcase the scalability and generalization of our proposed method across different image resolutions.

We evaluated the proposed model using a classifier with a six-layer convolution [66] on the STL-10 dataset, which consists of 5,000 labeled training data and 100,000 unlabeled training data. Figure 17 shows the confusion matrix of the classifier's predictions on the 8,000 test data. Similar to [38, 67], we followed the base papers bad Gan [51] and ICT [46] and adopted many of their hyper-parameters. We empirically set the coefficients of the consistency losses as $\lambda_1 = 5$, $\lambda_2 = 10$, $\lambda_3 = 0.5$. The confusion matrix includes the detection values 732, 724, 643, 748, 596, 671, 504, 693, 637, and 746 for the ten classes. Additionally, Fig. 18 showcases the images generated by the regularized bad generator, which supported the classifier. Furthermore, Fig. 19 showcases the convergence of the accuracy curve obtained from the model on the STL-10 data.

4.6 Vision transformer results

To further enhance the performance of our CRBSGAN framework, we investigated the integration of vision transformers, which have emerged as a powerful alternative to convolutional neural networks (CNNs) in computer vision [68]. In our study, we incorporated the vision transformer method into the discriminator of CRSSGAN to leverage its ability to capture long-range dependencies and model global image context. Our main objective was to improve the discriminating power and feature representation of the model by replacing specific components of the discriminator with vision transformers.

In our experiment, we replaced the original discriminator with a vision transformer-based discriminator. The generator and classifier remained consistent throughout the entire experiment. The discriminator transformer was configured with a patch size of 4, three input channels, and a single output class. The vision transformer itself had a hidden dimension of 384 and four attention heads [69]. During the training process, we utilized the gradient penalty loss and estimated the accuracy of the STl-10 data. Due to the limitations of our 8 GB GPU memory, we had to limit the batch size to 8 to ensure smooth execution.

Figures 20 and 21 depict the results, including the confusion matrices and generated images, obtained using a discriminator with/without a transformer on STL-10 images after 10 epochs. The model performance and quality of the generated images improved significantly with the application of the transformer. Additionally, our model, via the discriminator transformer after 10 epochs with a batch size of 8, reduced the error rate from 0.35 to 0.29. This integration allowed us to harness the attention mechanisms and self-attention mechanisms of vision transformers, resulting in a more comprehensive and informative signal being provided to the generator.

5 Discussion

5.1 Quantitative discussion

In order to demonstrate the effectiveness and originality of our proposed method, we conducted a comparison with SOTA baseline approaches. This comparison was initially conducted on the MNIST dataset, which was chosen due to its simplicity in comparison to other databases such as CIFAR-10 and SVHN. Additionally, we ensured a fair comparison by using uncomplicated and identical network architectures. To establish the superiority of our proposed approach, we compared it with existing methods based on bad generators, which we considered to be the SOTA methods for the MNIST dataset.

The CRBSGAN model was developed by incorporating local image augmentation, consistency regularization, and adversarial training into the bad generator, discriminator, and classifier. The architecture of the discriminator, bad generator, and classifier used for the MNIST data was based on the margin GAN model [51], which was used as the base model for comparison. However, the base model did not apply local image augmentation or consistency regularization. Table 2 presents the mean error rate percentages achieved using the CRBSGAN method on the MNIST dataset with 100, 600, 1000, and 3000 labeled training images, which were 2.99, 2.46, 2.35, and 1.56, respectively. In comparison, the base model achieved error rates of 3.53, 3.03, 2.87, and 2.06 with the same number of images [51]. The proposed semi-supervised CRBSGAN model outperforms the basic margin model [51], the CCS-GAN model [50], and the ICT method

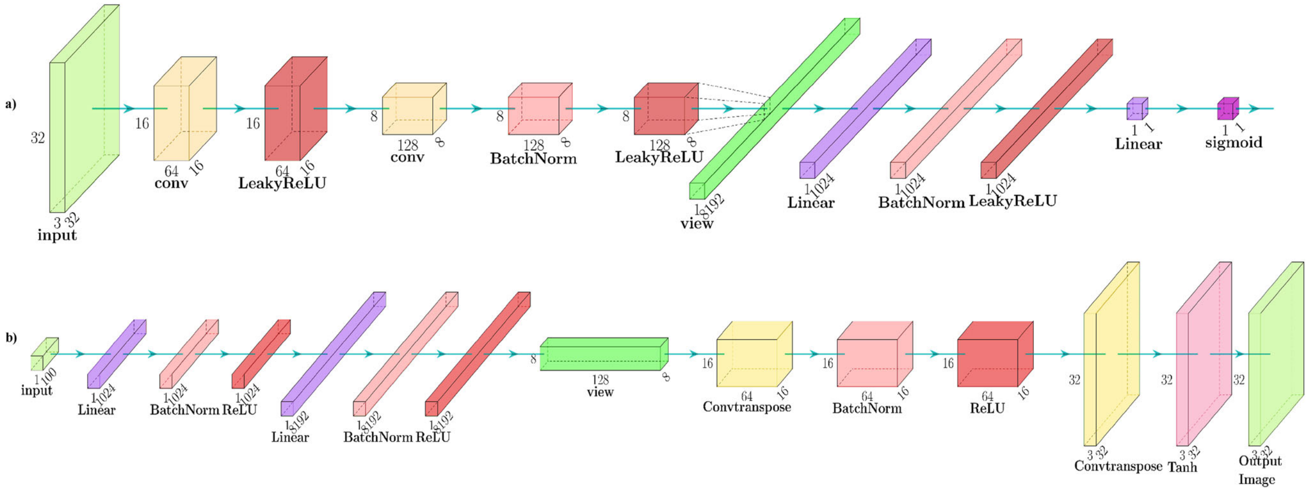


Fig. 10 The architecture of the **a** discriminator, **b** generator for SVHN, CIFAR-10

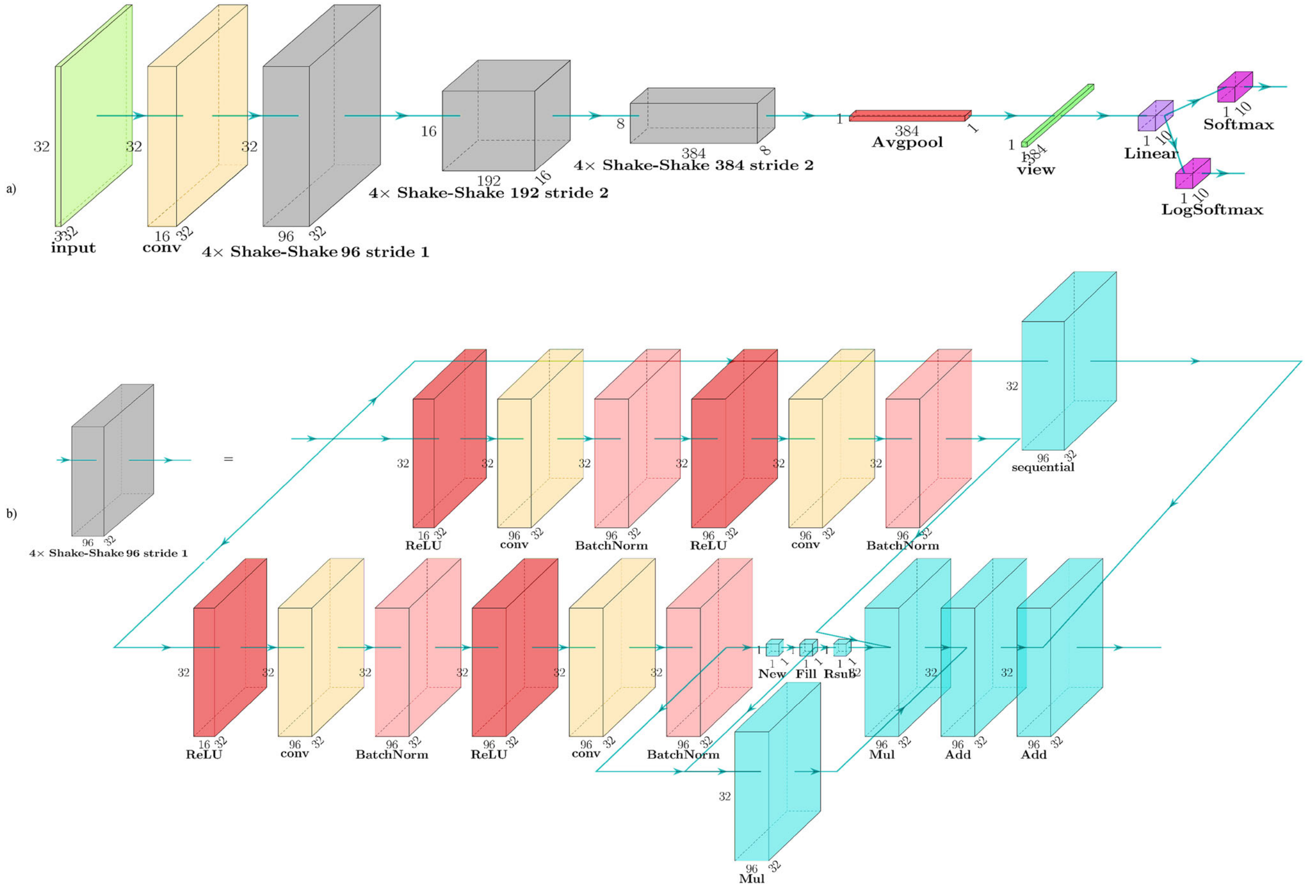


Fig. 11 Architecture of **a** the classifier for SVHN and CIFAR-10, **b** Shake-Shake 96 block details

[46], which used identical discriminator, generator, and classifier network parameters, as demonstrated in Table 2. The improvement results from regularizing the bad generator's fake informative and augmented images. By enlarging the images in Fig. 9, it is evident that the class of the samples can only be determined with a low degree of certainty. This observation suggests that the samples exhibit shared features between multiple classes. With the help of the visual insights from these images near the decision boundary, it was possible to increase the accuracy of the classifier and improve pseudo-labeling for unlabeled samples.

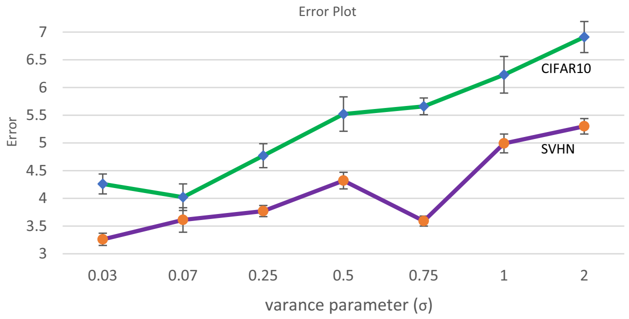


Fig. 12 Error plot for the CIFAR-10 and SVHN datasets

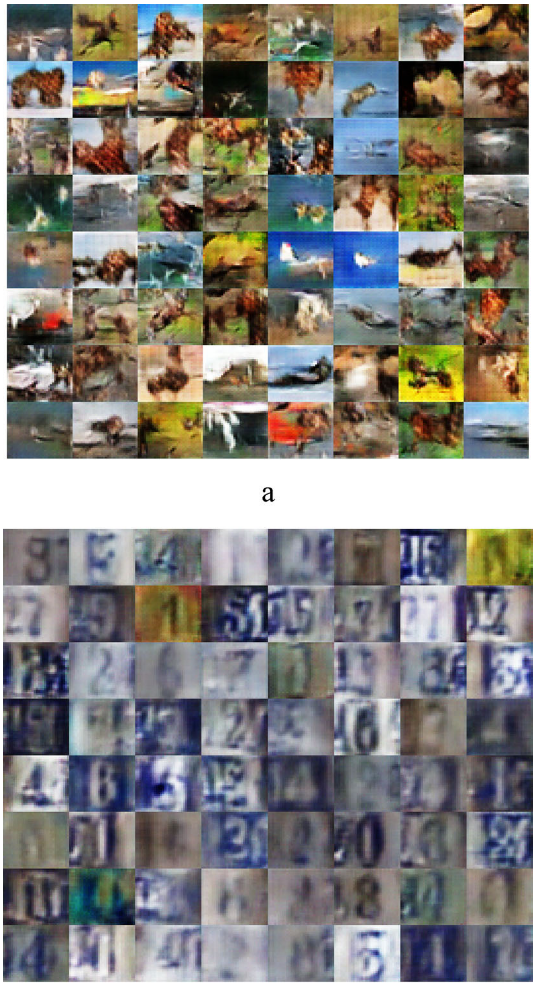
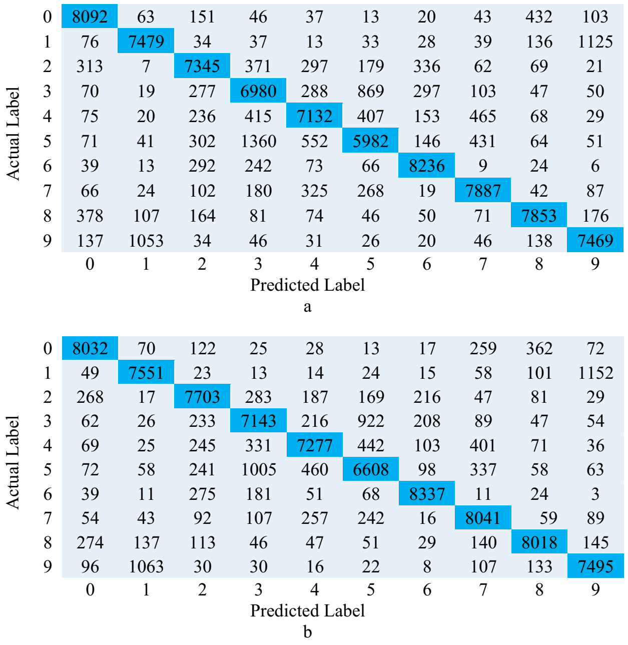


Fig. 13 Generated fake images by bad generators for **a** CIFAR-10 and **b** SVHN

To compare the effectiveness of our proposed method with existing approaches on the CIFAR-10 and SVHN datasets, we used bad GANs and the most recent good GANs with the same neural network architecture (shake-shake). Compared to the MNIST dataset, these datasets are more complex.



Actual Label	0	1	2	3	4	5	6	7	8	9
0	8092	63	151	46	37	13	20	43	432	103
1	76	7479	34	37	13	33	28	39	136	1125
2	313	7	7345	371	297	179	336	62	69	21
3	70	19	277	6980	288	869	297	103	47	50
4	75	20	236	415	7132	407	153	465	68	29
5	71	41	302	1360	552	5982	146	431	64	51
6	39	13	292	242	73	66	8236	9	24	6
7	66	24	102	180	325	268	19	7887	42	87
8	378	107	164	81	74	46	50	71	7853	176
9	137	1053	34	46	31	26	20	46	138	7469

Predicted Label

a

Actual Label	0	1	2	3	4	5	6	7	8	9
0	8032	70	122	25	28	13	17	259	362	72
1	49	7551	23	13	14	24	15	58	101	1152
2	268	17	7703	283	187	169	216	47	81	29
3	62	26	233	7143	216	922	208	89	47	54
4	69	25	245	331	7277	442	103	401	71	36
5	72	58	241	1005	460	6608	98	337	58	63
6	39	11	275	181	51	68	8337	11	24	3
7	54	43	92	107	257	242	16	8041	59	89
8	274	137	113	46	47	51	29	140	8018	145
9	96	1063	30	30	16	22	8	107	133	7495

Predicted Label

b

Fig. 14 Confusion matrices for classifiers trained with **a** and **b** 700 and 1000 labeled images per class on the CINIC-10 test images

Table 3 presents the mean error rates of the test images on the CIFAR-10 dataset with 1000 and 4000 labeled training samples using our proposed CRBSGAN method, which were 7.62% and 4.02%, respectively. These results were obtained using the same number of labeled images and conditions as the base model (10.39% and 6.44%), as reported in [51]. In contrast, the Triple-GAN-v2 (shake-shake) [74] and the AFDA model [75] achieved error rates of 8.41% and 6.05%, respectively.

For the SVHN dataset with 500 labeled training samples, our proposed method improved the mean error rate by 3.26% compared to the base model's error rate of 6.07% [51] and outperformed the Triple-GAN-v2's error rate of 3.61% [74]. Our proposed method improves performance by using local image augmentation, consistency regularization, and adversarial training to boost the bad generator.

Table 4 presents a comprehensive comparative analysis of the performance results on the CINIC-10 dataset. It compares the proposed approach with the currently available SOTA semi-supervised learning algorithms such as ICT [46], DSSLDDR + MT [40], and DNLL [41] methods. The semi-supervised ICT method [46], leveraging unlabeled images, achieved the following error rates: 25.81 ± 0.16 and 23.19 ± 0.21. The DSSLDDR + MT method [40] exhibited error rates of 23.96 ± 0.42 and 21.81 ± 0.16, while the DNLL method [41] resulted in error rates of 22.11 ± 0.28 and 19.38 ± 0.17. Furthermore, the CRBSGAN method proposed in this study demonstrated enhancements in the predicted error rates. It achieved values of 17.28 ± 0.19 and 15.32 ± 0.14

Fig. 15 Fake images generated by bad generators **a** and **b** on the CINIC-10 dataset where classifiers were trained with 700, and 1000 labeled images per class, respectively

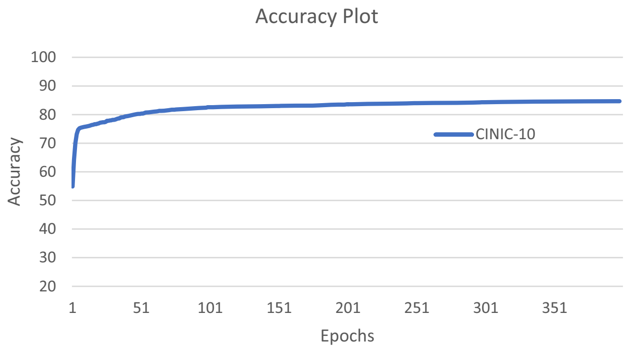
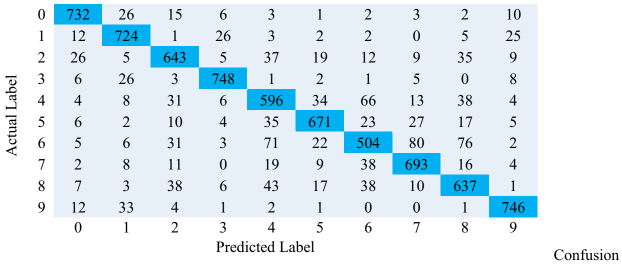


Fig. 16 The accuracy curve on the CINIC-10 data



	0	1	2	3	4	5	6	7	8	9
0	732	26	15	6	3	1	2	3	2	10
1	12	724	1	26	3	2	2	0	5	25
2	26	5	643	5	37	19	12	9	35	9
3	6	26	3	748	1	2	1	5	0	8
4	4	8	31	6	596	34	66	13	38	4
5	6	2	10	4	35	671	23	27	17	5
6	5	6	31	3	71	22	504	80	76	2
7	2	8	11	0	19	9	38	693	16	4
8	7	3	38	6	43	17	38	10	637	1
9	12	33	4	1	2	1	0	0	1	746

Actual Label
Predicted Label
Confusion

Fig. 17 Confusion matrix for the classifier trained with 500 labeled images per class on the STL-10 test images

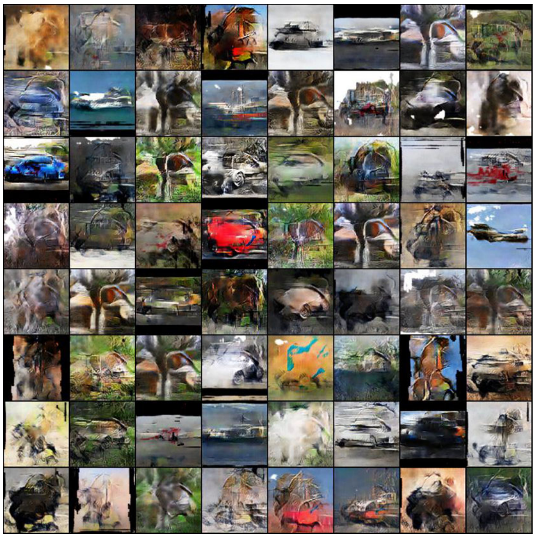


Fig. 18 Fake images generated by the bad generator on the STL-10 dataset

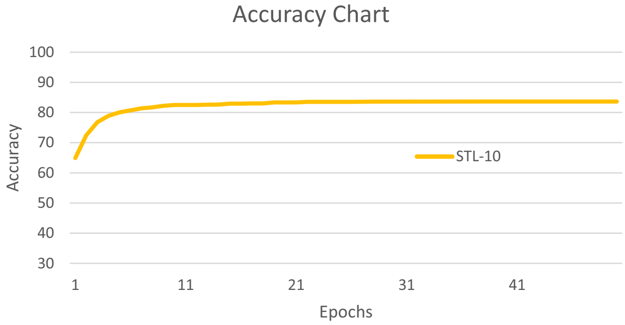
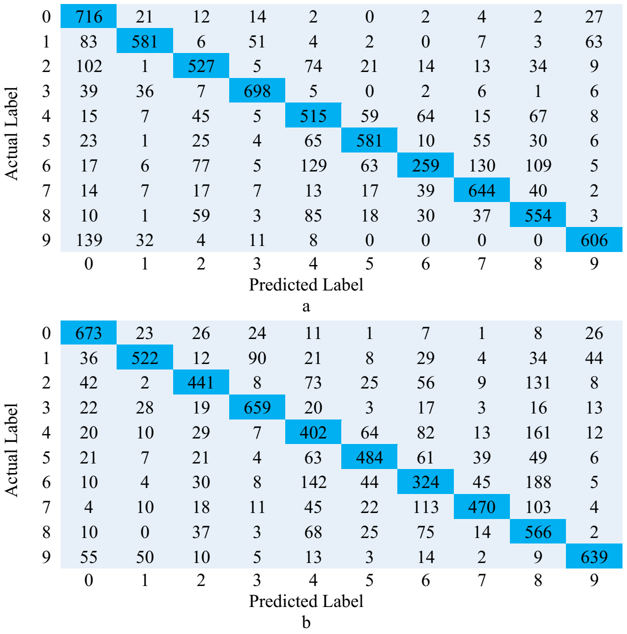


Fig. 19 The accuracy curve on the STL-10 data

using the CNN-13 classifier and 700 and 1000 labels per class, respectively.

In Table 5, a comparative analysis of the model performance, including several semi-supervised models, on the STL-10 test data is presented. The CNN-6 layer classifier, trained using 5,000 labeled images, achieved an error rate of 29.3. On the other hand, CRBSGAN with the assistance of a bad generator and discriminator using unlabeled data, resulted in error rates of 16.34 ± 0.07. As a result, this modification expanded the capabilities of our framework and facilitated a thorough comparison and evaluation of the advantages and trade-offs associated with utilizing transformers within the CRBSGAN framework.


Actual Label
Predicted Label
a
Actual Label
Predicted Label
b

	0	1	2	3	4	5	6	7	8	9
0	716	21	12	14	2	0	2	4	2	27
1	83	581	6	51	4	2	0	7	3	63
2	102	1	527	5	74	21	14	13	34	9
3	39	36	7	698	5	0	2	6	1	6
4	15	7	45	5	515	59	64	15	67	8
5	23	1	25	4	65	581	10	55	30	6
6	17	6	77	5	129	63	259	130	109	5
7	14	7	17	7	13	17	39	644	40	2
8	10	1	59	3	85	18	30	37	554	3
9	139	32	4	11	8	0	0	0	0	606

	0	1	2	3	4	5	6	7	8	9
0	673	23	26	24	11	1	7	1	8	26
1	36	522	12	90	21	8	29	4	34	44
2	42	2	441	8	73	25	56	9	131	8
3	22	28	19	659	20	3	17	3	16	13
4	20	10	29	7	402	64	82	13	161	12
5	21	7	21	4	63	484	61	39	49	6
6	10	4	30	8	142	44	324	45	188	5
7	4	10	18	11	45	22	113	470	103	4
8	10	0	37	3	68	25	75	14	566	2
9	55	50	10	5	13	3	14	2	9	639

Fig. 20 Confusion matrices for classifiers trained with a discriminator via/without transformer **a** and **b** on the STL-10 test images

5.2 Qualitative discussion

By studying the related works section, the main difference between non-generative adversarial network-based approaches and generative adversarial network-based approaches to semi-supervised classification is how they leverage unsupervised learning to improve classification performance. Non-generative adversarial network-based approaches typically rely on techniques such as self-training, co-training, and multi-view learning to utilize unlabeled images for semi-supervised classification [11]. These methods often involve training multiple classifiers on different subsets or views of the images and iteratively refining the classification boundaries based on the labeled and unlabeled images. However, good GANs refer to GANs that are well-trained and produce high-quality generated images that are visually similar to real images. The advantage of good generative adversarial network-based approaches is that they can potentially generate an unlimited number of synthetic samples, providing a rich source of additional training images for semi-supervised learning, while bad GANs can refer to GANs that have poor generators and produce low-quality generated images that have information about decision boundary.

The CRBSGAN method builds a new approach to semi-supervised learning, particularly in generating bad fake images as support vectors to reduce wrong pseudo-labeling. However, it introduces several novel elements that differentiate it from previous approaches and contribute to its improved performance.

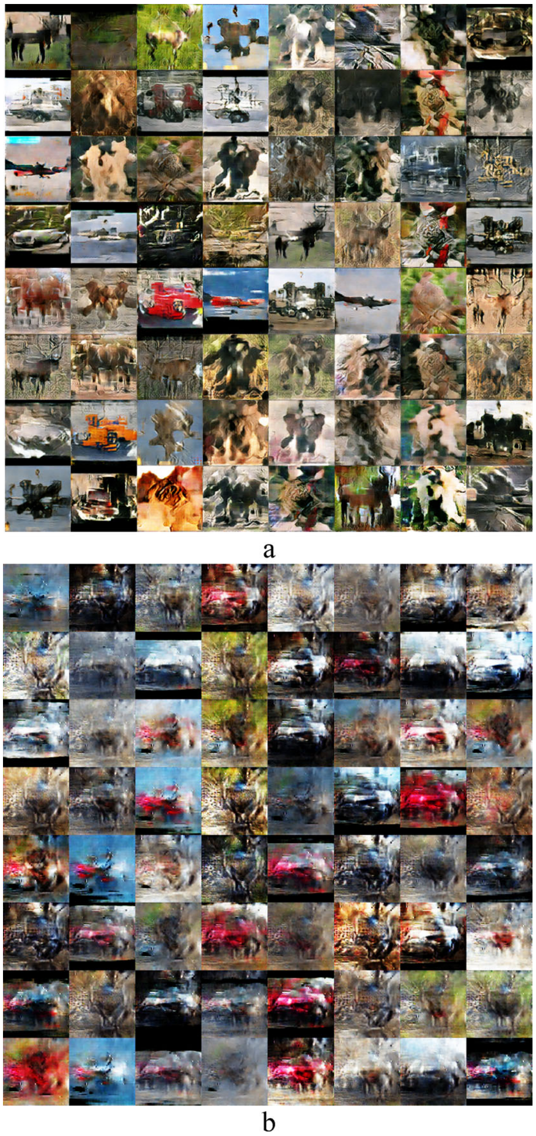


Fig. 21 Bad generated images via **a**/without **b** transformer on STL-10 data set after 10 epochs

1. One key difference is the use of local image augmentation, which generates more informative bad fake images near the decision boundary to improve the accuracy of pseudo-labels for low-confidence unlabeled images. This approach is more effective than previous methods that generate fake images uniformly across the feature space. By generating more informative bad fake images near the decision boundary, the CRBSGAN model can better capture the distribution of the underlying images and improve its ability to generalize to new images.
2. Another novel element is the use of consistency regularization, which encourages the model to produce similar outputs for perturbed versions of the input. It helps to reduce over-fitting and improve the generalization performance of the model. Previous methods have used

Table 2 Means and standard deviation (SD) of the error rates (%) for the MNIST test images (over five runs)

Reference	Model	Number of labels			
		100	600	1000	3000
[70]	NN	25.81	11.44	10.70	6.04
[70]	SVM	23.44	8.85	7.77	4.21
[70]	CNN	22.98	7.68	6.45	3.35
[70]	TSVM	16.81	6.16	5.38	3.45
[70]	EmbedNN	16.86	5.97	5.73	3.59
[71]	DBN-rNCA	–	8.70	–	3.30
[72]	CAE	13.47	6.30	4.77	3.22
[73]	MTC	12.03	5.13	3.64	2.57
[57]	dropNN	21.89	8.57	6.59	3.72
[57]	+ PL	16.15	5.03	4.30	2.80
[57]	+ PL + DAE	10.49	4.1	3.46	2.69
[51]	Base model bad GAN	3.53 ± 0.57	3.03 ± 0.6	2.87 ± 0.71	2.06 ± 0.2
[50]	CCS-GAN	8.34	3.81	3.37	2.49
[46]	ICT	6.96 ± 0.45	4.48 ± 0.02	3.34 ± 0.33	2.21 ± 0.09
Ours	Consistency- Regularized bad GAN	**2.99** ± **0.19**	**2.46** ± **0.25**	**2.35** ± **0.41**	**1.56** ± 0.23

The best results are highlighted by bolding the values

Table 3 Comparison of Means and SD of the error rate (%) for the SVHN and CIFAR10 test images (over five runs)

Reference	Model	Number Of labels		
		SVHN (500)	CIFAR-10(1000)	CIFAR-10(4000)
[76]	Ladder	–	–	20.04 ± 0.47
[77]	CatGAN	–	–	19.58 ± 0.58
[78]	FM GANs	18.44 ± 0.48	19.61 ± 0.20	18.63 ± 2.32
[48]	Triple-GAN	–	–	18.82 ± 0.32
[79]	SGAN	–	–	17.26 ± 0.69
[80]	π model	6.83 ± 0.66	27.36 ± 1.20	13.20 ± 0.27
[23]	Bad GAN (share layers)	6.20 ± 0.07	18.37 ± 0.55	14.5 ± 0.26
[51]	Base model bad Gan(3player)	6.07 ± 0.43	10.39 ± 0.43	6.44 ± 0.10
[75]	AFDA	–	9.40 ± 0.32	6.05 ± 0.13
[18]	VAT + Ent	–	–	10.55 ± 0.05
[46]	ICT	–	15.48 ± 0.78	7.29 ± 0.02
[50]	CCS-GAN	5.19 ±0.13	15.80 ± 0.22	14.01 ± 0.15
[74]	Triple-GAN-v2 (shake-shake)	3.61 ±0.26	8.41 ± 0.19	6.54 ± 0.08
Ours	Consistency-Regularized bad GAN (3player-shake-shake)	**3.26** ± **0.11**	**7.62** ± 0.35	**4.02** ± 0.24

The best results are highlighted by bolding the values

Table 4 Average error rate (%) for the CINIC-10 test data obtained from five runs

Reference	Model	Number Of labels	
		CINIC-10 (700)	CINIC-10 (1000)
[80]	π model	29.66 ± 1.12	27.04 ± 0.85
[35]	TE	30.38 ± 1.01	27.35 ± 0.86
[35]	MT	28.41 ± 0.29	25.71 ± 0.12
[46]	ICT	25.81 ± 0.16	23.19 ± 0.21
[40]	DSSLDDR	29.35 ± 0.31	26.75 ± 0.24
[40]	*DSSLDDR + MT*	23.96 ± 0.42	21.81 ± 0.16
[41]	*DNLL*	22.11 ± 0.28	19.38 ± 0.17
Ours	Consistency-Regularized bad GAN	**17.28 ± 0.19**	**15.32 ± 0.14**

The best results are highlighted by bolding the values

consistency regularization, but the CRBSGAN approach extends it by incorporating local image augmentation and adversarial training for bad generators further to improve the consistency of the model's outputs. The CRBSGAN approach also utilizes adversarial training, which involves training a discriminator to distinguish between bad fake and real images. It helps to improve the diversity and quality of the bad fake images, leading to better performance on the classification task.

3. The CRBSGAN model utilizes consistency regularization to boost the model's classifier and reduce the classifier's margin on pseudo-labels generated for fake images, which are used to train the model in adversarial training. This approach is more effective than previous methods in that the classifier's prediction in adversarial training was not reinforced. They had included noisy and irrelevant margins on pseudo-labels for bad fake images, which degraded the model's performance.
4. The CRBSGAN incorporated a transformer-based discriminator that enhances the performance of the bad generator through transformer attention in the three-player bad semi-supervised generative adversarial network framework.

These elements combine to improve the efficiency, accuracy, and robustness of the model, leading to significant improvements in classification performance compared to previous SOTA methods.

Table 5 Average error rate (%) for the STL-10 test data obtained from five runs

Reference	Model	STL-10 (5000)
[66]	CNN	29.3 ± NA
[66]	*CNN* + Adversarial attacks	25 ± NA
[46]	ICT	27 ± 0.18
[51]	bad GAN	28.2 ± 0.23
Ours	Consistency-Regularized bad GAN	**16.34 ± 0.07**

The best results are highlighted by bolding the values

5.3 Theoretical discussion

Suppose we have training images $S=\ (X, Y) = \{(x_i, y_i)|x_i \in \mathbb{R}^{d*d}, y_i \in \{1 \ldots K\}\}_{i=1}^{N}$ with $p_{X,Y}^{\text{real}}$ distribution. Real images (X, Y) are divided into labeled $X^L = (X^L, \text{Y}) \sim p_{(x^l, y)}^{\text{real}} = \left\{\left(x_i^l, y_i\right)|x_i^l \in \mathbb{R}^{d*d}, y_i \in \{1 \ldots K\}\right\}_{i=1}^{V}$ and unlabeled images $X^U \sim p_{x^u}^{\text{real}} = \left\{\left(x_i^u\right)|x_i^u \in \mathbb{R}^{d*d}\right\}_{i=1}^{Q}$where$V + Q = N, V \ll Q$, and fake images $X^G \sim p_{x^g}^{\text{fake}} = \left\{\left(x_i^g\right)|x_i^g \in \mathbb{R}^{d*d}\right\}_{i=1}^{M}$ are generated by a bad generator. The total images are set $T = S \cup X^G$. There may be many predictors $h : X \rightarrow Y$ that map input X to output Y in supervised classification. We are seeking a predictor $\widehat{h}$ that minimizes the empirical risk $R = \mathrm{E}_{X^L, Y}\mathbb{I}[Y \neq h(X^L)]$ on X^L (Eq. 17).

$$\widehat{h} = \underset{h}{argmin} R \tag{17}$$

In semi-supervised classification, the predictive model $\widehat{h}$ aims to minimize the empirical risk for less labeled images and more unlabeled images. The p function of the model $\widehat{h}$ provides a probability vector belonging to each class for unlabeled images $q^U = p\left(Y|X^U\right)$. The model $\widehat{h}$ aims to bring the class probability vector to the maximum probability class $\widehat{\text{Y}}$=arg max (q^U) closer via cross-entropy [57]. In reference [23], it is shown that a bad generator in the two-player game reduces wrong pseudo-labeling and empirical risk $\widehat{h}$ (Eq. 18).

$$\hat{h} = \underset{h}{\text{arg min}}\ \mathrm{E}_{X^L, Y}\mathbb{I}[Y \neq h(X)] + \mathrm{E}_{X^U} L_{\text{CE}}\left(q^U, \hat{Y}\right) \tag{18}$$

Labeled and unlabeled real images X^L, X^U are provided to the discriminator with label 1 and generated fake images $X^G$with label 0. A weak augmentation set B is generated for these images, and label consistency regularization is performed. Assume that set Bcontains invariant label augmentations $\alpha(T)$ on the labeled, unlabeled, and fake images $Tas\left[h(T) = h(\alpha(T)),\ \alpha \text{ is a tolerable augmentation function}\right]$ (Eqs.19–22). The proposed model's discriminator, generator, and classifier are deep neural networks. According to [55], the consistency-regularized discriminator $\ddot{h}_d$ shows less empirical risk and generalization error than the usual discriminator $\hat{h}_d$, which causes the production of more informative images through the consistency-regularized bad generator. We define an upper bound for the consistency-regularized discriminator $\ddot{h}_d$ and classifier $\ddot{h}_c$ based on [55], per Eqs. 23 and 24. These theoretical findings suggest that image augmentation and consistency regularization may aid in the improvement of bad generative adversarial networks.

$$\begin{aligned} X^{AL} &= \left(X^{AL}, Y\right) \sim p_{(x^l, y)}^{\text{aument - real}} = \left(x_i^{al}, y\right) \\ &= \Big\{\alpha\left(x_i^l, y_i\right)|\left(x_i^l, y_i\right) \in (X^L, Y),\ \alpha \text{ is weak} \\ &\quad \text{augmentation function, } h\left(X^L\right) = h\left(\alpha\left(X^L\right)\right)\Big\}_{i=1}^{O*V}, \\ &\quad O \text{ is augmentation factor} \end{aligned} \tag{19}$$

$$\begin{aligned} \mathrm{X}^{AU} &\sim p_{x^u}^{\text{aument - real}} = \left(x_i^{au}\right) \\ &= \Big\{\alpha\left(x_i^u\right)|x_i^u \in X^U,\ \alpha \text{ is weak augmentation function,} \\ &\quad h\left(X^U\right) = h\left(\alpha\left(X^U\right)\right)\Big\}_{i=1}^{O*Q} \end{aligned} \tag{20}$$

$$\begin{aligned} \mathrm{X}^{AG} &\sim p_{x^g}^{\text{aument - fake}} = \left(x_i^{ag}\right) \\ &= \Big\{\alpha\left(x_i^g\right)|x_i^g \in X^G,\ \alpha \text{ is weak augmentation function,} \\ &\quad h\left(X^G\right) = h\left(\alpha\left(X^G\right)\right)\Big\}_{i=1}^{O*M} \end{aligned} \tag{21}$$

$$\mathrm{B} = \alpha(T) = X^{AL} \cup X^{AU} \cup X^{AG} \tag{22}$$

$$R\left(\ddot{h}_d\right) - R\left(\hat{h}_d\right) \leq \sqrt{\frac{2\log 2 + \log\left(\frac{1}{\delta}\right)}{N}} \tag{23}$$

$$R\left(\ddot{h}_c\right) - R\left(\hat{h}_c\right) \leq \sqrt{\frac{K\log K + \log\left(\frac{1}{\delta}\right)}{N}} \tag{24}$$

6 Conclusion

Data scarcity is detrimental to supervised machine learning. Correct pseudo-labeling can enhance the classifier's performance when leveraging unlabeled images. This research aimed to address the problem of incorrect pseudo-labeling of unlabeled images using a novel three-player framework termed CRBSGAN achieved through a new loss function. The proposed model includes bad generators that produce low-quality images, which contain information about the decision boundary. A bad fake image augmentation and a good image augmentation better covered the data space in bad GAN. Additionally, a novel consistency-regularized bad generator was developed using the new consistency regularization of bad fake images. The discriminator, bad generator, and classifier components were strengthened by proposed consistency regularizations. Also, replacing the transformer-based discriminator with a pure discriminator improved the generation of bad images. This study demonstrated that the consistency-regularized bad semi-supervised GAN is effective at pseudo-labeling for unlabeled images, and the proposed model outperformed previous research in terms of error rate. By improving the classification performance using unlabeled images, our research contributes to the development of AI models that can better understand and analyze visual information.

The proposed approach has various applications in computer-generated imagery (CGI) and virtual worlds. It enables diverse content generation, improves visual analysis, and enhances the fidelity and consistency of style transfer or artistic rendering algorithms in CGI and virtual worlds. Additionally, it results in more accurate identification and tracking of objects, precise segmentation, realistic lighting effects, and enhanced visual aesthetics. These applications demonstrate the versatility and potential impact of the suggested method in these domains.

Using advanced architectures of deep networks, such as changing network depth and layer type in the generator, discriminator, and classifier, may yield better results in future research. Additionally, the consistency regularization of deep network weights for the generator, discriminator, and classifier will almost certainly produce impressive results. Furthermore, the performance of the proposed model could be improved by adjusting hyper-parameters such as the initial weights of deep network layers, the size of convolution filters, and the coefficients of losses using meta-heuristic algorithms. This approach may lead to improved results and better adaptability of the model across different datasets. Another potential avenue is to explore integrating Mobile-Sal's efficient feature extraction capabilities, particularly leveraging depth information, into a bad GAN architecture to enhance the quality and diversity of generated samples [81]. One of the study's limitations is unbalanced data processing, which can be investigated using custom loss functions or data balancing methods. In instances where high-quality unlabeled images are limited, employing a good generator to produce such data is anticipated to enhance model efficiency.

Authors' contribution Iraji and Tanha proposed the Consistency-Regularized Bad Semi-Supervised Generative Adversarial Networks approach. Iraji executed the approach and analyzed the results. Iraji, Tanha, Balafar, and Feizi-Derakhshi were responsible for the manuscript's conceptualization, validation, resources, and editing. All authors read and authorized the final manuscript.

Data availability Data will be made available on request.

Declarations

Competing interest The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Ethical and informed consent This article does not contain any studies with human participants or animals performed by any of the authors. The datasets used in the manuscript are derived from publicly available data sets and may be obtained from the appropriate authors upon reasonable request.

References

1. Qin, Y., et al.: GuideRender: large-scale scene navigation based on multi-modal view frustum movement prediction. Vis. Comput. **39**(8), 3597–3607 (2023)
2. Sheng, B., et al.: Accelerated robust Boolean operations based on hybrid representations. Comput. Aided Geom. Des. **62**, 133–153 (2018)
3. Jiang, J., et al.: Real-time hair simulation with heptadiagonal decomposition on mass spring system. Graph. Models **111**, 101077 (2020)
4. Ertugrul, E., et al.: Embedding 3D models in offline physical environments. Comput. Anim. Virtual Worlds **31**(4–5), e1959 (2020)
5. Huo, X., et al.: Attention regularized semi-supervised learning with class-ambiguous data for image classification. Pattern Recogn. **129**, 108727 (2022)
6. Jian, C., Yang, K., Ao, Y.: Industrial fault diagnosis based on active learning and semi-supervised learning using small training set. Eng. Appl. Artif. Intell. **104**, 104365 (2021)
7. Chang, J.-H., Weng, H.-C.: Fully used reliable data and attention consistency for semi-supervised learning. Knowl.-Based Syst. **249**, 108837 (2022)
8. Ren, Q., et al.: A framework of active learning and semi-supervised learning for lithology identification based on improved naive Bayes. Expert Syst. Appl. **202**, 117278 (2022)
9. Gu, X.: A self-training hierarchical prototype-based approach for semi-supervised classification. Inf. Sci. **535**, 204–224 (2020)
10. Lu, L., et al.: Uncertainty-aware pseudo-label and consistency for semi-supervised medical image segmentation. Biomed. Signal Process. Control **79**, 104203 (2023)
11. Zhang, Y., et al.: Multi-view classification with semi-supervised learning for SAR target recognition. Signal Process. **183**, 108030 (2021)
12. Emadi, M., et al.: A selection metric for semi-supervised learning based on neighborhood construction. Inf. Process. Manage. **58**(2), 102444 (2021)
13. Wei, X., et al.: FMixCutMatch for semi-supervised deep learning. Neural Netw. **133**, 166–176 (2021)
14. Zhang, B., et al.: Flexmatch: Boosting semi-supervised learning with curriculum pseudo labeling. Adv. Neural. Inf. Process. Syst. **34**, 18408–18419 (2021)
15. Arantes, R.B., Vogiatzis, G., Faria, D.R.: Learning an augmentation strategy for sparse datasets. Image Vis. Comput. **117**, 104338 (2022)
16. Xiu, Y., et al.: FreMix: Frequency-based mixup for data augmentation. Wirel. Commun. Mob. Comput. **2022** (2022)
17. Gan, Y., et al.: Deep semi-supervised learning with contrastive learning and partial label propagation for image data. Knowl.-Based Syst. **245**, 108602 (2022)
18. Miyato, T., et al.: Virtual adversarial training: a regularization method for supervised and semi-supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. **41**(8), 1979–1993 (2018)
19. Gangwar, A., et al.: Triple-BigGAN: Semi-supervised generative adversarial networks for image synthesis and classification on sexual facial expression recognition. Neurocomputing **528**, 200–216 (2023)
20. He, R., et al.: Generative adversarial network-based semi-supervised learning for real-time risk warning of process industries. Expert Syst. Appl. **150**, 113244 (2020)
21. Liu, Y., et al.: Regularizing discriminative capability of CGANs for semi-supervised generative learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (2020)
22. Li, Y., et al.: The theoretical research of generative adversarial networks: an overview. Neurocomputing **435**, 26–41 (2021)
23. Dai, Z., et al.: Good semi-supervised learning that requires a bad gan. Adv, Neural Inf. Process. Syst. **30** (2017)
24. Yun, S., et al.: Cutmix: Regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF international conference on computer vision. (2019)
25. Goodfellow, I., et al.: Generative adversarial networks. Commun. ACM **63**(11), 139–144 (2020)
26. Wang, R., et al.: Better pseudo-label: Joint domain-aware label and dual-classifier for semi-supervised domain generalization. Pattern Recogn. **133**, 108987 (2023)
27. Kim, D., et al.: Multi-co-training for document classification using various document representations: TF–IDF, LDA, and Doc2Vec. Inf. Sci. **477**, 15–29 (2019)
28. Yu, K., et al.: A consistency regularization based semi-supervised learning approach for intelligent fault diagnosis of rolling bearing. Measurement **165**, 107987 (2020)
29. Liu, L., Tan, R.T.: Certainty driven consistency loss on multi-teacher networks for semi-supervised learning. Pattern Recogn. **120**, 108140 (2021)
30. Ke, Z., et al.: Dual student: Breaking the limits of the teacher in semi-supervised learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. (2019)
31. Deng, W., et al.: Deep ladder reconstruction-classification network for unsupervised domain adaptation. Pattern Recogn. Lett. **152**, 398–405 (2021)
32. Xiao, H., et al.: Semi-supervised semantic segmentation with cross teacher training. Neurocomputing **508**, 36–46 (2022)
33. Li, B., Pi, D., Lin, Y.: Learning ladder neural networks for semi-supervised node classification in social network. Expert Syst. Appl. **165**, 113957 (2021)
34. Chen, J., Yang, M., Ling, J.: Attention-based label consistency for semi-supervised deep learning based image classification. Neurocomputing **453**, 731–741 (2021)
35. Meel, P., Vishwakarma, D.K.: A temporal ensembling based semi-supervised ConvNet for the detection of fake news articles. Expert Syst. Appl. **177**, 115002 (2021)

36. Ding, W., Abdel-Basset, M., Hawash, H.: RCTE: A reliable and consistent temporal-ensembling framework for semi-supervised segmentation of COVID-19 lesions. Inf. Sci. **578**, 559–573 (2021)
37. Wang, J., et al.: Adversarial attacks and defenses in deep learning for image recognition: A survey. Neurocomputing **514**, 162–181 (2022)
38. Berthelot, D., et al.: Remixmatch: Semi-supervised learning with distribution alignment and augmentation anchoring. Int. Conf. Learn. Represent. (ICLR), (2020)
39. Sohn, K., et al.: Fixmatch: Simplifying semi-supervised learning with consistency and confidence. Adv. Neural. Inf. Process. Syst. **33**, 596–608 (2020)
40. Yang, M., et al.: Discriminative semi-supervised learning via deep and dictionary representation for image classification. Pattern Recogn. **140**, 109521 (2023)
41. Xu, H., et al.: Semi-supervised learning with pseudo-negative labels for image classification. Knowl.-Based Syst. **260**, 110166 (2023)
42. Li, X., et al.: Feature-aware conditional GAN for category text generation. Neurocomputing **547**, 126352 (2023)
43. Rubin, M., et al.: TOP-GAN: Stain-free cancer cell classification using deep learning with a small training set. Med. Image Anal. **57**, 176–185 (2019)
44. Mao, J., et al.: Pseudo-labeling generative adversarial networks for medical image classification. Comput. Biol. Med. **147**, 105729 (2022)
45. Chen, Z., Ramachandra, B., Vatsavai, R.R.: Consistency regularization with generative adversarial networks for semi-supervised learning (2020). arXiv preprint arXiv:2007.03844
46. Verma, V., et al.: Interpolation consistency training for semi-supervised learning. Neural Netw. **145**, 90–106 (2022)
47. Zhao, Z. et al.: Improved consistency regularization for gans. In: Proceedings of the AAAI Conference on Artificial Intelligence (2021)
48. Li, C. et al.: Triple generative adversarial nets. Adv. Neural Inf. Process. Syst. **30** (2017)
49. Gan, Y. et al.: Generative adversarial networks with adaptive learning strategy for noise-to-image synthesis. Neural Comput. Appl. **35**(8), 6197–6206 (2022)
50. Wang, L., Sun, Y., Wang, Z.: CCS-GAN: A semi-supervised generative adversarial network for image classification. Vis. Comput. **38**(6), 2009–2021 (2022)
51. Dong, J., Lin, T.: MarginGAN: Adversarial training in semi-supervised learning. Adv. Neural Inf. Process. Syst. **32** (2019)
52. Gu, X., Angelov, P.P.: Semi-supervised deep rule-based approach for image classification. Appl. Soft Comput. **68**, 53–68 (2018)
53. Zhang, H. et al.: Consistency regularization for generative adversarial networks. Proc. Int. Conf. Learn. Represent. (2020)
54. Yang, M., et al.: Deep neural networks with L1 and L2 regularization for high dimensional corporate credit risk prediction. Expert Syst. Appl. **213**, 118873 (2023)
55. Yang, S. et al.: Sample efficiency of data augmentation consistency regularization. In: International Conference on Artificial Intelligence and Statistics. PMLR (2023)
56. Feng, W., et al.: New margin-based subsampling iterative technique in modified random forests for classification. Knowl.-Based Syst. **182**, 104845 (2019)
57. Lee, D.-H.: Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks. In: Workshop on challenges in representation learning, ICML. (2013)
58. Liu, Z., et al.: Dual-feature-embeddings-based semi-supervised learning for cognitive engagement classification in online course discussions. Knowl.-Based Syst. **259**, 110053 (2023)
59. Li, W., et al.: Tackling mode collapse in multi-generator GANs with orthogonal vectors. Pattern Recogn. **110**, 107646 (2021)
60. LeCun, Y., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)
61. Netzer, Y. et al.: Reading digits in natural images with unsupervised feature learning. In: NIPS workshop on deep learning and unsupervised feature learning. 2011, Granada, Spain.
62. Darlow, L.N. et al.: Cinic-10 is not imagenet or cifar-10 (2018). arXiv preprint arXiv:1810.03505
63. Krizhevsky, A., Hinton, G.: Learning multiple layers of features from tiny images (2009)
64. Coates, A., Ng, A., Lee, H.: An analysis of single-layer networks in unsupervised feature learning. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics. JMLR Workshop and Conference Proceedings. (2011)
65. Qiu, S., et al.: Adversarial attack and defense technologies in natural language processing: A survey. Neurocomputing **492**, 278–307 (2022)
66. Zoppi, T., Ceccarelli, A.: Detect adversarial attacks against deep neural networks with GPU monitoring. IEEE Access **9**, 150579–150591 (2021)
67. Bao, J. et al.: CVAE-GAN: fine-grained image generation through asymmetric training. In: Proceedings of the IEEE international conference on computer vision. (2017)
68. Wu, Y.-H. et al.: P2T: Pyramid pooling transformer for scene understanding. IEEE Trans. Pattern Anal. Mach. Intell. (2022)
69. Jiang, Y., Chang, S., Wang, Z.: Transgan: Two pure transformers can make one strong gan, and that can scale up. Adv. Neural. Inf. Process. Syst. **34**, 14745–14758 (2021)
70. Weston, J., Ratle, F., Collobert, R.: Deep learning via semi-supervised embedding. In: Proceedings of the 25th international conference on Machine learning. (2008)
71. Salakhutdinov, R., Hinton, G.: Learning a nonlinear embedding by preserving class neighbourhood structure. In: Artificial Intelligence and Statistics. PMLR (2007)
72. Ranzato, M.A. et al.: Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: 2007 IEEE conference on computer vision and pattern recognition. IEEE (2007)
73. Rifai, S. et al.: The manifold tangent classifier. Adv. Neural Inf. Process. Syst. **24** (2011)
74. Li, C., et al.: Triple generative adversarial networks. IEEE Trans. Pattern Anal. Mach. Intell. **44**(12), 9629–9640 (2021)
75. Mayer, C., Paul, M., Timofte, R.: Adversarial feature distribution alignment for semi-supervised learning. Comput. Vis. Image Underst. **202**, 103109 (2021)
76. Rasmus, A. et al.: Semi-supervised learning with ladder networks. Adv. Neural Inf. Process. Syst. **28** (2015)
77. Springenberg, J.T.: Unsupervised and semi-supervised learning with categorical generative adversarial networks. Proceedings of International Conference on Learning Representations (ICLR), (2016)
78. Salimans, T. et al.: Improved techniques for training gans. Adv. Neural Inf. Process. Syst. **29** (2016)
79. Deng, Z. et al.: Structured generative adversarial networks. Adv. Neural Inf. Process. Syst. **30** (2017)
80. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Adv. Neural Inf. Process. Syst. **30** (2017)
81. Wu, Y.-H., et al.: MobileSal: Extremely efficient RGB-D salient object detection. IEEE Trans. Pattern Anal. Mach. Intell. **44**(12), 10261–10269 (2021)

Mohammad Saber Iraji is a faculty member of the Computer Science Department at PNU University in Iran. His research areas include semi-supervised classification, pattern recognition, image processing, feature selection, fuzzy logic, and game theory.

Jafar Tanha obtained his Bachelor's degree in Computer Science in December 1999 and his Master's degree in Computer Science and Applied Mathematics in June 2002 from the Department of Computer Science and Mathematics of the University of Amir Kabir (Polytechnic Tehran). He received a Ph.D. degree in the Computer Science Department from the University of Amsterdam. He is currently a Professor at the Faculty of Computer Engineering, University of Tabriz, Iran. His research interests include learning (artificial intelligence), pattern classification, semi-supervised learning, pattern clustering, complex networks, and game theory.

Mohammad-Ali Balafar is currently a Professor at the Faculty of Computer Engineering, University of Tabriz, Iran. His research interests include feature extraction, biomedical MRI, brain, convolutional neural nets, face recognition, fuzzy set theory, image classification, image enhancement, image motion analysis, image representation, image resolution, image segmentation, image sensors, image sequences, image texture, medical image processing, Bayes methods, probability, and wavelet transforms.

Mohammad-Reza Feizi-Derakhshi received a B.S. degree in software engineering from the University of Isfahan, Iran, and an M.Sc. and Ph.D. degree in artificial intelligence from the Iran University of Science and Technology, Tehran, Iran. He is currently a Professor at the Faculty of Computer Engineering, University of Tabriz, Iran. His research interests include natural language processing, optimization algorithms, deep learning, social network analysis, and intelligent databases.