

Received 1 May 2023, accepted 13 May 2023, date of publication 16 May 2023, date of current version 24 May 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3276875

RESEARCH ARTICLE

Low-Sample Image Classification Based on Intrinsic Consistency Loss and Uncertainty Weighting Method

ZHIGUO LI¹, LINGBO LI², XI XIAO³, JINPENG CHEN⁴,
NAWEI ZHANG⁵, AND SAI LI⁶, (Member, IEEE)

¹Information Construction and Service Center, Neijiang Normal University, Neijiang 641100, China

²Library of Information Center, Zhejiang Technical Institute of Economics, Hangzhou 310018, China

³School of Computer Science, Southwest Petroleum University, Chengdu 610000, China

⁴Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115, USA

⁵College of Information Science and Engineering, China University of Petroleum, Beijing 102249, China

⁶College of Mechanical and Electrical Engineering, Zaozhuang University, Zaozhuang 277160, China

Corresponding author: Sai Li (lisaizxy@163.com)


This work was supported in part by the Natural Science Foundation of Shandong Province under Grant ZR202103050458, and in part by the Startup Foundation for Doctoral Research of Zaozhuang University under Grant 1020714.

ABSTRACT As is well known, the classification performance of large deep neural networks is closely related to the amount of annotated data. However, in practical applications, the quantity of annotated data is minimal for many computer vision tasks, which poses a considerable challenge for deep convolutional neural networks that aim to achieve ideal classification performance. This paper proposes a new, fully supervised low-sample image classification model to alleviate the problem of limited marked sample quantity in real life. Specifically, this paper presents a new sample intrinsic consistency loss, which can more effectively update model parameters from a “fundamental” perspective by exploring the difference between intrinsic sample features and semantic information contained in sample labels. Secondly, a new uncertainty weighting method is proposed to weigh the original supervised loss. It can more effectively learn sample features by weighting sample losses one by one based on their classification status and help the model autonomously understand the importance of different local information. Finally, a sample generation model generates some artificial samples to supplement the limited quantity of actual training samples. The model adjusts parameters through the combined effect of sample intrinsic consistency loss and weighted supervised loss. This paper uses 25 % of the SVHN dataset and 30 % of the CIFAR-10 dataset as training samples to simulate scenarios with limited sample quantities in real life, achieving accuracies of 94.59 % and 91.27 % respectively, demonstrating the effectiveness of our method on small real datasets.

INDEX TERMS Low-sample image classification, deep convolutional neural network, sample intrinsic consistency loss, uncertainty weighting method, image generation model.

I. INTRODUCTION

The initial research on image classification [1] relied on manually identifying image features such as shape, color, and texture to classify images accurately. Today, deep neural networks (DNNs) have gained widespread attention for their ability to map low-level features to higher levels and

The associate editor coordinating the review of this manuscript and approving it for publication was Mingbo Zhao .

abstract out higher-level features to discover the underlying distribution patterns of training samples. Through continuous research in recent years, large-scale deep neural networks [2], [3], [4], [5], [6] have acquired more powerful function representation and feature extraction capabilities and have made significant breakthroughs in real-world applications such as image classification [7], image segmentation [8], and object detection [9]. However, these successes are not only due to the continuous improvement of deep learning methods but also

to the abundant availability of well-labeled training data for network training, which is even more crucial.

Researchers have proposed many classification models [7], [10], [11], [12], [13], [14] for autonomous image classification, aiming to reduce the labor and additional cost required for image annotation work [15], [16], [17], [18]. It is well known that the performance of these deep neural networks is closely related to the number of labeled samples. However, annotated data are often scarce for some classification tasks in daily life. In some cases, collecting data is even more difficult than annotating data [19], which leads to small sample datasets becoming the norm in everyday life and limits the effectiveness of large neural networks for these tasks. Therefore, this paper focuses on addressing the problem of the severely limited amount of annotated data in fully supervised classification tasks.

In recent years, image generation models have been a hot topic for their ability to create artificial images that resemble actual samples by learning the underlying distribution of the data. These generated dummy images improve the image classification performance of the model by complementing the smaller number of real examples in the model training process. Nowadays, Generative Adversarial Networks (GAN) [20] and Variational Autoencoders (VAEs) [21] are almost the dominant image generation models. VAEs generate never-before-seen images by learning the underlying distribution of samples and mapping it to image space. It consists of two parts, an encoder and a decoder, and this encoder-decoder framework has been widely used for image generation. Recently, many proposed methods have utilized the principle of Variational Autoencoders (VAEs) to generate images [22], [23], [24]. Generative adversarial networks consist of generators and discriminators. The generator generates “convincing” fake photos, while the discriminator predicts the probability that the generated images come from the entire training set. As the generator and discriminator compete, the generator eventually produces images that resemble the training samples. Unsupervised and semi-supervised learning has achieved unexpected results by using generative adversarial networks and many variants based on them [25], [26], [27], [28] to generate some artificial images resembling actual examples to complement the original pictures.

In some specific computer image classification tasks in real life, only a small amount of labeled data is available for model training, which may affect the classification performance of large neural networks. Therefore, this article proposes a fully supervised low-sample classification model to solve this problem. Specifically, a new sample intrinsic consistency loss (SICL) is proposed to explore the difference between sample label connotation information and intrinsic sample features. Secondly, a new uncertainty weighting method (UW) is proposed, which can precisely weight the loss of each sample based on its classification situation. Finally, the image generation model generates some artificial samples to supplement the limited labeled samples. Our main contributions are as follows.

- 1) This paper proposes a new low-sample image classification model that can achieve good classification results even with a limited number of labeled samples. It effectively alleviates the problem of deep neural network classification performance degradation caused by insufficient labeled data in real-world scenarios.
- 2) A new sample intrinsic consistency loss is proposed in this paper, which can effectively update model parameters by exploring the differences between sample label connotation information and intrinsic sample features.
- 3) This paper also proposes a new uncertainty weighting method, which can precisely weight each sample loss according to its classification situation, aiming to make the model understand the importance of different local features of the sample.
- 4) In addition, a new image generation model is proposed based on ACGAN, which can generate high-quality labeled samples more effectively and expand the size of the dataset.

II. RELATED WORK

Although image classification methods based on deep learning differ, they all adopt some basic common principles, which can also be understood as technical means in algorithm training, including the type of loss function and image generation methods. This section introduces some common basic strategies and their principles in image classification methods and proposes preliminary improvement strategies for the areas we believe need modification.

A. SAMPLE SEMANTIC RELATIONSHIPS

Recently, researchers have increasingly focused on the abstract features of image samples inside the network. Li et al. [29] proposed a new data enhancement method to cope with the problem of limited data by studying the internal components of sample networks to improve the classification performance of EEG processing. López et al. [30] improved the overfitting network phenomenon caused by unbalanced data by exploring the intrinsic features of the data. Mantoo and Khurana [31] proposed an Android malware detection system based on inherent data features, achieving 97.5 % accuracy.

Many semi-supervised models have also achieved good image classification performance by studying the semantic information between different samples within the network [32], [33]. They enforce the consistency of the semantic information of two sub-images obtained from an unlabeled image in the network to learn the sample features. However, this enforced consistency needs a valuable metric to measure it. Therefore, this article takes the actual label mapping of the network internal samples as the “metric,” compares it with the sample features extracted from the network internal samples, and uses it as the sample intrinsic consistency loss.

For example, now give three students a test paper to do together. Previously, semi-supervised image classification studied the semantic information of samples within the

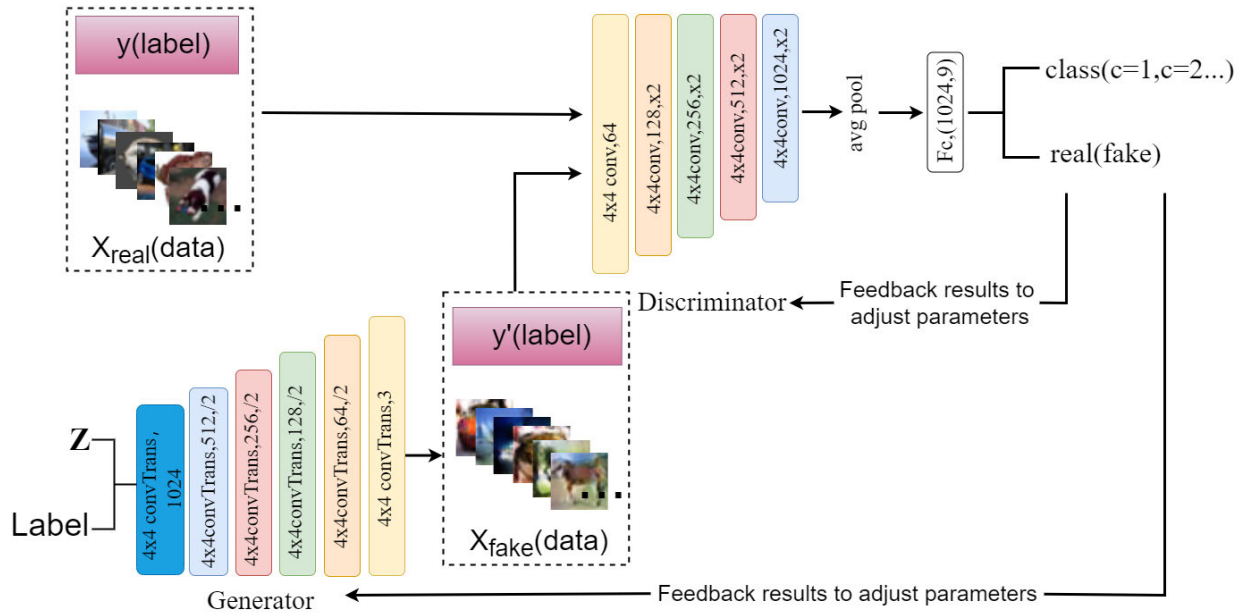


FIGURE 1. The general framework for auxiliary classifier generative adversarial network (ACGAN).

network, just like two students discussing their answers after completing the test paper to discover mistakes and correct them. In contrast, the sample intrinsic consistency loss is like the third student correcting his problems by comparing them with the answers of the test paper itself. So it seems the sample intrinsic consistency loss is more “explicit.”

B. MODEL TRAINING LOSS

In existing classification models, the predicted sample values are always made to better fit the actual sample labels by minimizing the cross-entropy loss [34]. The cross-entropy loss is defined as follows.

$$L_C = -\frac{1}{\text{batch_size}} \sum_{j=1}^{\text{batch_size}} \sum_{i=1}^n y_{ji} \log y'_{ji} \quad (1)$$

Here, batch_size and n represent the minibatch containing B samples and the total number of image categories, respectively. y_{ji} and y'_{ji} represent the actual label of the sample and the predicted sample label, respectively, and L_C represents the difference between the existing label and the predicted label of the sample. However, the cross-entropy function only makes relatively general statistics on the differences between all samples' actual and predicted labels. This may lead to misclassified samples needing to be classified into the correct category in future model training.

Xu et al. [35] derive a semantic loss function that bridges the gap between the neural output vector and the logical constraints. This loss function captures how close the neural network is to satisfying its output constraints. Improving on the softmax loss function, Maharjan et al. [36] proposed a brain tumor detection scheme with an accuracy improvement of nearly 2% and a processing time reduction of 50 ms.

Chen et al. [37] proposed a correlated entropy-induced loss function (CLF) to improve model performance and experimentally demonstrated that CLF can make deep learning models more robust.

Focal loss [38] and reduced focal loss [39] make the model pay more attention to the difficulty distinguishing features of training samples, which undoubtedly plays an important role. However, using uncertainty-weighted methods to weight the original supervised loss, it is possible to adjust the loss of each sample according to its specific classification situation. The weighted supervised loss is more targeted, and the model parameters can be updated more effectively during the training process behind the model.

C. AUXILIARY CLASSIFIER GENERATIVE ADVERSARIAL NETWORK (ACGAN)

Generative adversarial networks [20] (GAN) consist of a generator and a discriminator. During the “co-growth” of the generator and discriminator, it generates fake images that resemble the actual training samples (without labels). The auxiliary classifier generative adversarial network (ACGAN) [28], on the other hand, can independently generate artificial images (carrying brands) similar to the actual training samples. This solves the extra labeling work required because the GAN generates unlabeled samples [20], as shown in Fig 1. The numbers of the networks in the generator and discriminator indicate the size of the convolution kernel, the number of sample feature channels, and the step size of the convolution, respectively. Specifically, it first integrates the information of randomly generated labels and random noise Z into the generator to generate some counterfeit ideas. These phony images resemble the underlying distribution of training

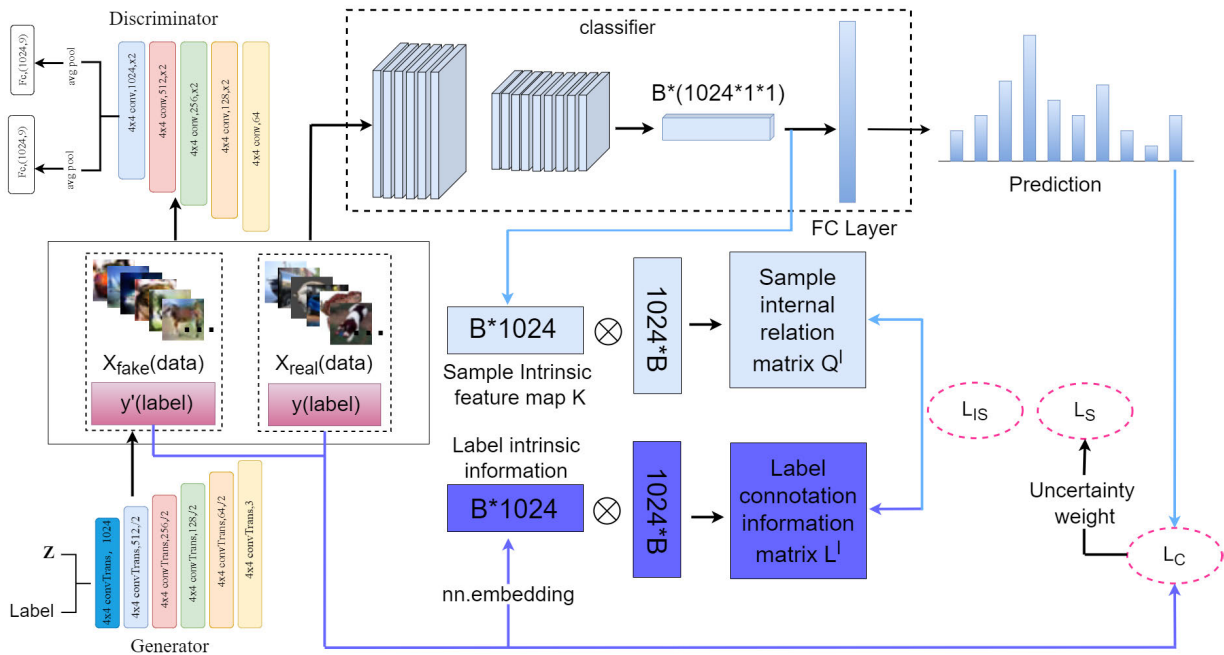


FIGURE 2. The general framework of our low-sample image classification model.

images and carry brands. Then, these fake images $X_{fake}(data)$, and authentic images $X_{real}(data)$ are fed to the discriminator simultaneously. The discriminator needs not only to recognize the authenticity of the pictures but also to classify them. Finally, with the competition between the generator and the discriminator, the generator produces a labeled false sample $X_{fake}(data)$ similar to the actual training sample.

$$L_S = E [\log P(S = real | X_{real}) + \log P(S = fake | X_{fake})] \quad (2)$$

$$L_T = E [\log P(C = c | X_{real}) + \log P(C = c | X_{fake})] \quad (3)$$

where L_S and L_T represent the discriminator’s ability to recognize image authenticity and correctly classify all data, respectively, the generator is trained to maximize $L_T - L_S$ (the parts of L_S and L_T about the actual image $X_{real}(data)$ are independent of the generator), i.e., the generator wants to maximize L_T and minimize L_S . To make the generated information more realistic, the generator tries to reduce the probability that its generated data $X_{fake}(data)$ will be discriminated as false. At the same time, the generator wants to maximize the likelihood that the generated data will be correctly classified. The discriminator is trained to maximize $L_T + L_S$, which maximizes its ability to organize and identify true and false data accurately.

D. SHARED DISCRIMINATOR ARCHITECTURE

The output of a traditional classifier for image classification using adversarial generative networks is a $k+1$ probability distribution [28], [40], [41]. Like ACGAN [28], the discriminator consists of two “head” or final layers, one for

image category classification and the other for distinguishing between true and false images. The k th+1st output of the discriminator indicates the probability of whether the image is true or false. However, combining two tasks (distinguishing true from false and classifying) may degrade the performance of both functions [19]. The authors of Triple Generative Adversarial Networks [42] claim that if the discriminator contains two incompatible tasks: image classification and discriminating true from false, then the performance of both functions is degraded. In recent years, many approaches using adversarial generative networks for image classification have started utilizing individual networks with separate classification and discriminative branches [43], [44], [45]. Therefore, this paper will build a new independent network branch structure for image classification based on ACGAN, which can ensure that the two tasks of identifying the authenticity of images and classifying images do not conflict.

III. PROPOSED METHODS

Fig 2 shows the general framework of our proposed image classification model for the low-sample dataset, which consists of a generator (G), a discriminator (D), and a classifier.

During the formal training of the model, the image generation model first generates labeled artificial samples $X_{fake}(data)$ in batches (i.e., in sets during the training process) to supplement the actual number of restricted samples $X_{real}(data)$. The synthetic examples may be fuzzy or even a random combination of pixel points during the initial training process. Still, as the training process iterates, the generated fake images become more and more perfect. Then, we feed

all the images into the classifier and adjust our model parameters under the combined effect of the intrinsic consistency loss of the samples and the supervised loss after uncertainty weighting. This eventually allows our model to achieve relatively good classification results despite the limited number of pieces. During the iterative process, each small batch of generated fake images $X_{\text{fake}}(\text{data})$ is immediately input to the classifier, so the generated images are not saved.

A. SAMPLE INTRINSIC CONSISTENCY LOSS (SICL)

Both authors [32] and [33] argue that the sample features within the network contain more semantic information, and they achieve good results in their respective models by exploring this information. But they both learn sample features by comparing the differences between the intrinsic characteristics of two sub-pictures of the same unlabeled sample obtained by random perturbation. We correct the inherent qualities of the example by mapping the label information carried by the example to the internal network as the “standard” so that we can “target” and learn the sample features more fully and accurately.

We use a case-level Gram Matrix to represent the structured relation among various samples. Assuming a mini-batch with B samples, we denote the activation map of layer L as $F^l \in \mathbb{R}^{B \times C \times H \times W}$, where H and W are the spatial dimensions of the feature map, and C is the number of channels. We reshape the feature map F^l to $K^l \in \mathbb{R}^{B \times CHW}$ and then compute the Case-wise Gram Matrix G^l as follows:

$$G^l = K^l \cdot (K^l)^T \quad (4)$$

The similarity between the activations of the i -th and j -th samples in the input mini-batch is represented by the inner product of the vectorized activation map $K_{(i)}^l$ and $K_{(j)}^l$, denoted by G_{ij} . We obtain the sample relation matrix Q^l by applying L2 normalization to each row G_i^l of the Case-wise Gram Matrix G^l .

$$Q^l = \left[\frac{G_1^l}{\|G_1^l\|_2}, \dots, \frac{G_B^l}{\|G_B^l\|_2} \right]^T \quad (5)$$

Similarly, we use the label intrinsic relationship matrix L^l to represent the intrinsic information similarity of the labels corresponding to these B samples.

$$L^l = \left[\frac{L_1^l}{\|L_1^l\|_2}, \dots, \frac{L_B^l}{\|L_B^l\|_2} \right]^T \quad (6)$$

Sample intrinsic relationship loss requires the sample intrinsic semantic information to be consistent with the sample label connotation information to ensure the semantic relationship between samples. We define the proposed sample intrinsic relationship loss as follows:

$$L_{IS} = \sum_{x \in \{X_{\text{fake}}, X_{\text{real}}\}} \frac{1}{B} \|Q^l - L^l\|_2^2 \quad (7)$$

where x is the actual samples from the training set and some artificial samples generated, by minimizing L_{IS} during the training process, the network is enhanced to capture more robust and discriminative sample features, which helps to extract additional semantic information. The feature map obtained from the deeper layer contains more advanced information compared to the one obtained from the middle layer. As a result, the feature map before the final average pooling layer is utilized to compute the intrinsic feature matrix G^l for the sample.

B. UNCERTAINTY-WEIGHTED LOSS (UW)

The original supervised loss [34] calculates only the difference between the actual and predicted labels of the samples and then narrows their differences during the subsequent training of the model so that the predicted labels of the examples are consistent with their actual labels. Focused loss [38] makes the model more focused on the features of complex samples by weighting the original supervised loss but does not determine whether the sample-by-sample classification is correct. In contrast, our model can judge its classification results sample by sample and adjust our model parameters for the specific classification of each sample, as shown in Fig 3.

First, the actual training samples $X_{\text{real}}(\text{data})$ and the generated artificial samples $X_{\text{fake}}(\text{data})$ are fed into the classifier to obtain the sample prediction values. Then the probability corresponding to the sample prediction value is compared with the pre-set threshold (note that this threshold constantly changes).

$$P_{t_{\max}} \geq T \quad (8)$$

here $P_{t_{\max}}$ represents the probability corresponding to the sample prediction label, and T is our pre-set threshold value. Because the model's performance gets more robust as the number of training rounds increases, the threshold we set increases as the number of training rounds increases.

$$T = (1/N) + \left(\tau - \frac{1}{N} \right) / \text{epoch} * \text{epochs} \quad (9)$$

here the epoch and epochs represent the number of current training rounds and the total number of training rounds, respectively. N is the total number of sample categories; refer to [33], we set τ to 1. If the probability of the sample prediction value is less than this threshold, a smaller weight is directly assigned to this sample loss front. Conversely, proceed to the subsequent conditional judgment. The sample predicted value is compared with the accurate sample label, and a smaller weight is assigned if it is consistent, and a more considerable weight is assigned vice versa.

$$P_{\max} = P_{\text{real}} \quad (10)$$

here P_{\max} and P_{real} represent the image prediction labels and the actual sample labels, respectively. The weights are assigned to the expressions as follows.

$$\text{Weight} = \begin{cases} L_W = (1 - P_{t_{\max}})^2 \\ H_W = 1 \end{cases} \quad (11)$$

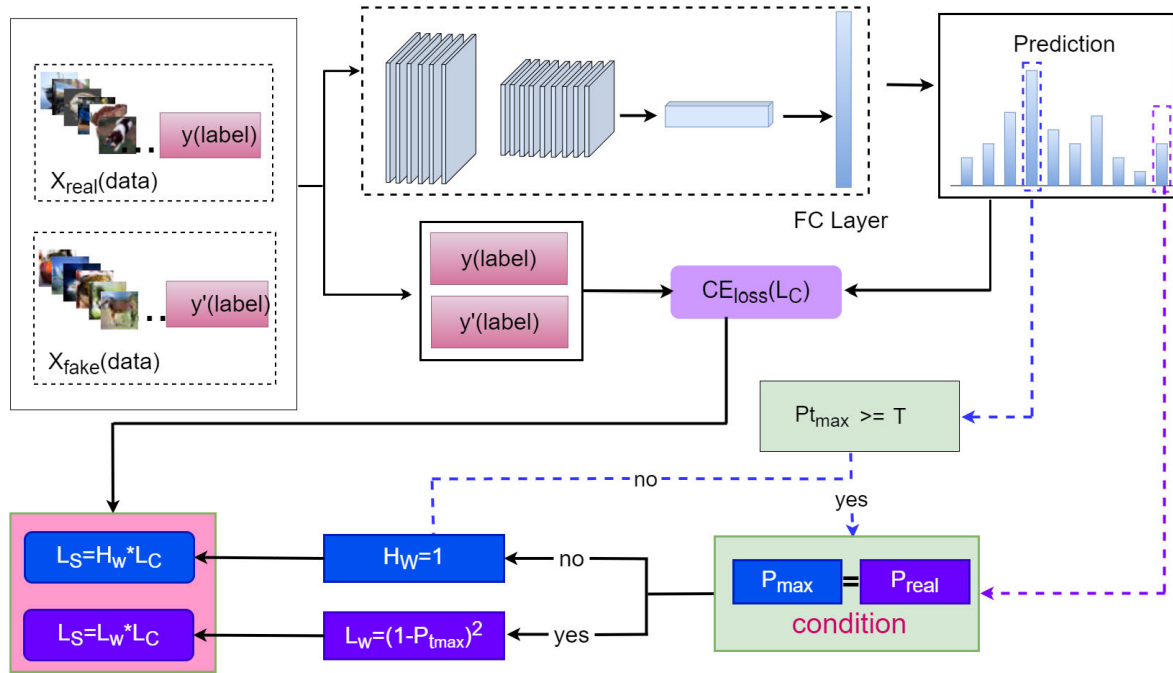


FIGURE 3. The framework for using uncertainty weighting methods for original supervisory.

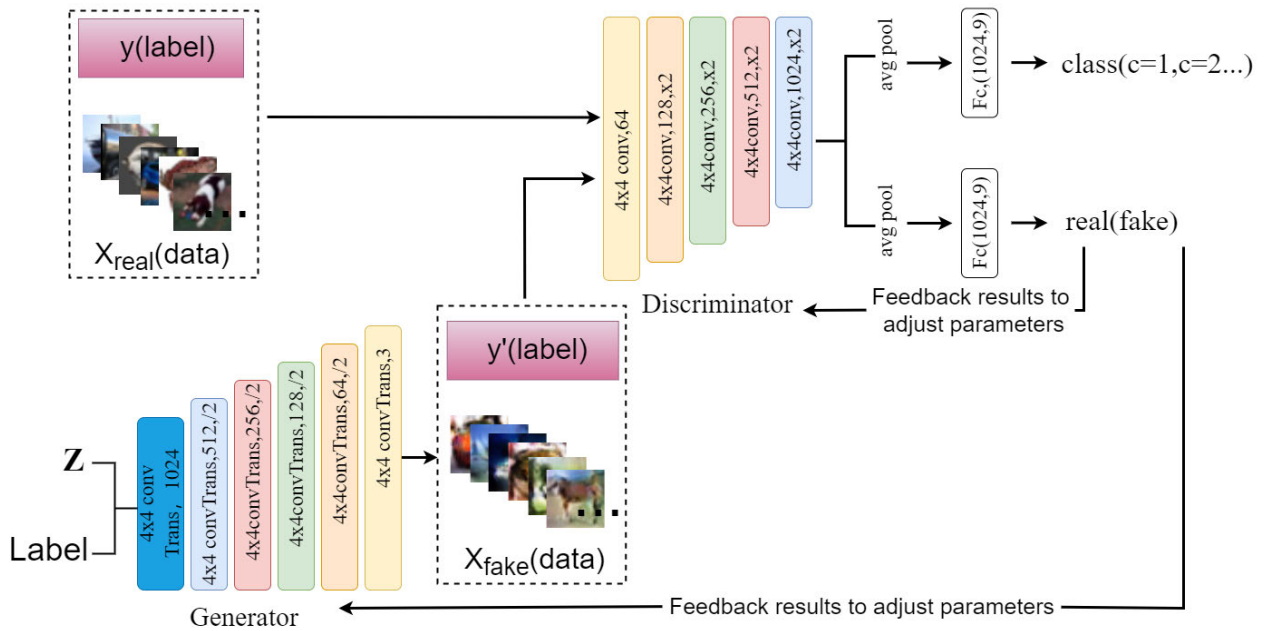


FIGURE 4. The general framework of the ACGAN-based image generation model.

The uncertainty-weighted loss ensures that the model can thoroughly learn their sample characteristics sample by sample and update the model parameters more efficiently. Its formula is as follows.

$$L_S = \begin{cases} L_w \times L_C & \text{if } P_{\max} = P_{\text{real}} \text{ and } P_{t_{\max}} \geq T \\ H_w \times L_C & \text{otherwise} \end{cases} \quad (12)$$

here L_C is the original supervised loss, L_S and L_{IS} are the uncertainty-weighted loss and the sample intrinsic consistency loss, respectively.

C. IMAGE GENERATION MODEL

As mentioned earlier, the performance of both functions is degraded if the discriminator is given two tasks

TABLE 1. The accuracy of our method compared with previous approaches on the test set using different proportions of SVHN dataset as training samples.

Dataset Size	Ours		EC-GAN		Shared ResNet Discriminator		Shared DC Discriminator	
	ResNet	ResNet+ACGAN	ResNet	ResNet+GAN	ResNet	ResNet+ACGAN	DCNet	DCNet+ACGAN
	5%	84.82	89.15	84.82	87.61	84.82	86.74	79.86
10%	88.45	91.38	88.45	90.79	88.45	89.31	83.52	86.13
15%	90.88	92.92	90.88	91.99	90.88	91.47	85.23	88.36
20%	92.64	93.87	92.64	93.35	92.64	93.13	86.85	89.39
25%	92.87	94.59	92.87	93.72	92.87	93.54	87.59	90.17

TABLE 2. The effect of the SICL and UW methods in training models using different percentages of SVHN datasets.

Method	5%	10%	15%	20%	25%
Classifier(ResNet)	84.82	88.45	90.88	92.64	92.87
Classifier(ResNet)+SICL	85.67	89.17	91.53	92.87	93.71
Classifier(ResNet)+UW	85.89	89.23	91.38	93.53	93.87
Ours	89.15	91.38	92.92	93.81	94.59

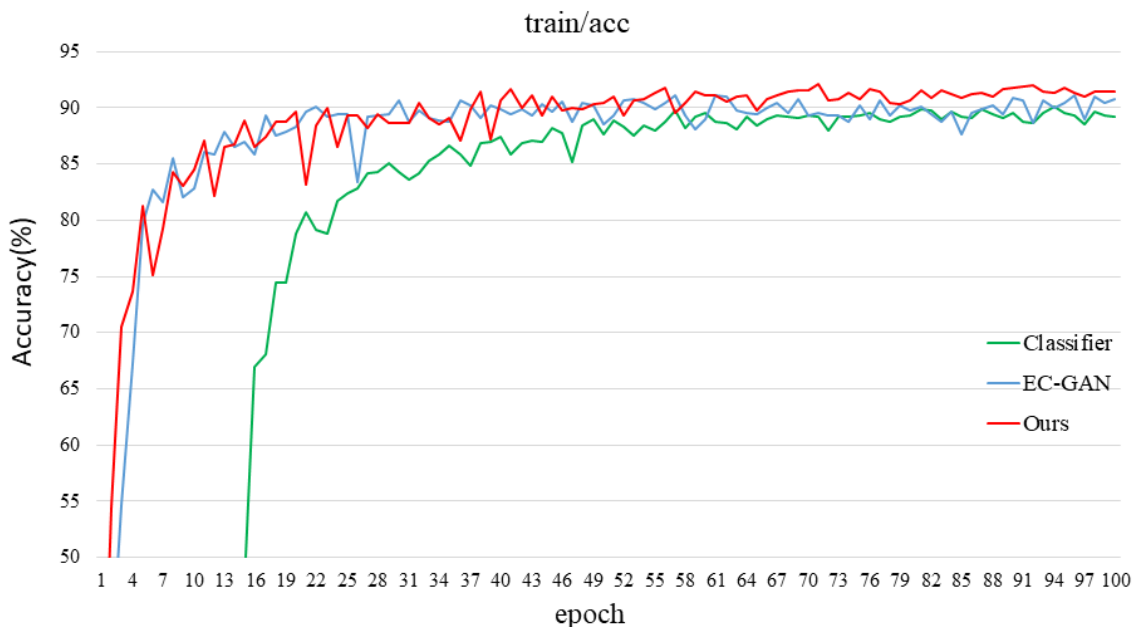


FIGURE 5. The accuracy of our method compared with the classifier and EC-GAN on the test set using 10% of the SVHN dataset as the training sample for the model.

simultaneously, i.e., image classification and determining image authenticity [19], [42]. Therefore, we created a new independent network branch based on ACGAN [28] to classify the generated and authentic images. As shown in Fig 4, this model separates the two tasks of image classification and judging image authenticity, which ensures that the generated samples match more closely with the underlying distribu-

tion of the actual samples and classify the training samples more accurately. The loss function (D) of the discriminator is defined as follows:

$$L(D) = BCE(D(X_{fake}), 0) + BCE(D(X_{real}), 1) \quad (13)$$

TABLE 3. The accuracy of our method compared with previous approaches on the test set using different proportions of CIFAR-10 dataset as training samples.

Dataset Size	Ours		EC-GAN		Shared ResNet Discriminator		Shared DC Discriminator	
	ResNet	ResNet+ACGAN	ResNet	ResNet+GAN	ResNet	ResNet+ACGAN	DCNet	DCNet+ACGAN
	10%	78.23	82.17	78.23	80.31	78.23	79.85	73.43
15%	81.16	85.42	81.16	83.75	81.16	82.17	76.82	77.58
20%	86.32	88.95	86.32	87.47	86.32	87.51	78.47	80.35
25%	88.27	90.37	88.27	89.52	88.27	89.35	81.06	82.16
30%	89.92	91.27	89.92	90.79	89.92	90.54	83.79	85.47

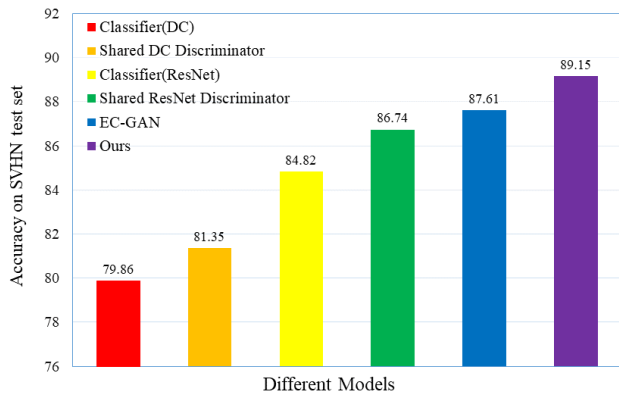


FIGURE 6. Using a 5% SVHN dataset as the model's training sample, our method's accuracy performance is compared with previous methods.

The loss function of the generator(G) is defined as:

$$L(G) = \text{BCE} (D (X_{\text{fake}}), 1) \tag{14}$$

The independent network branch loss function for the classification of images is:

$$L(C) = \text{CE} (C (X_{\text{fake}}), y') + \text{CE} (C (X_{\text{real}}), y) \tag{15}$$

where X_{real} and y represent the actual training samples with their labels, respectively, Z and y' (the labels required by the generator (G) to generate the fake example X_{fake} (data) are the random noise and randomly generated labels. $G(Z)$ represents the generated artificial sample X_{fake} (data) (data). BCE is the binary cross-entropy, and CE is the cross-entropy [27].

D. TOTAL MODEL TRAINING LOSS

Therefore, the total loss of our low-sample image classification model is.

$$L = (L_S (X_{\text{real}}) + L_{IS} (X_{\text{real}})) + \lambda (L_S (X_{\text{fake}}) + L_{IS} (X_{\text{fake}})) \tag{16}$$

Specifically, the first half of the formula is the overall loss of the actual training samples, and the second half is the widespread loss of the artificial samples. λ is the coefficient

that balances the loss of the authentic samples and the loss of the synthetic samples. X_{real} and X_{fake} are the actual training samples and the artificial samples (carrying labels) generated during model training. L_S is the weighted supervisory loss after weighting the original supervisory loss using the uncertainty weighting method, and L_{IS} is the sample intrinsic consistency loss.

IV. EXPERIMENTS AND ANALYSIS

In this section, we train our model with different proportions of samples from two large datasets (CIFAR-10 dataset and SVHN dataset) to simulate actual constrained samples in real scenarios and evaluate our approach based on the experimental results.

A. EXPERIMENTAL PARAMETER SETTINGS AND DETAILS

As mentioned, EC-GAN also tries to address the problem of poor model performance caused by small sample datasets. Therefore, our code is modified from EC-GAN and implemented in PyTorch, using a learning rate of 0.0002 [26], a normalized value of 0.5, and an Adam variant of the SGD optimizer [46] in the algorithm. We set the parameter λ in Equation 12 to 0.1 concerning EC-GAN [19]. For image enhancement, we use a random crop of 4×4 and a random rotation of 10 degrees, and an L2 regularization of 0.001 [47] or weight decay. During training, we manipulate the number of samples involved in model training by adjusting the dataset's percentage size to evaluate our method's effectiveness on small sample datasets.

This paper compares our approach with three previous models, EC-GAN, Shared ResNet Discriminator, and Shared DC Discriminator. Where our model and EC-GAN, Shared ResNet Discriminator both use the ResNet-18 neural network [14] as the classifier in the left column, and Shared DC Discriminator uses the deep convolutional neural network as the classifier. The right column of each method shows the classification results after adding various image generation models. It is worth noting that Shared ResNet Discriminator and Shared DC Discriminator use the original ACGAN as the image generation model.

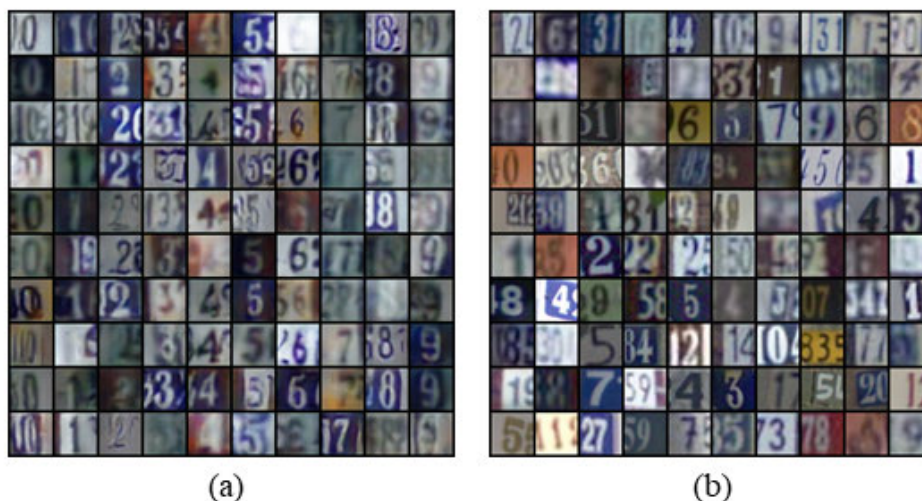


FIGURE 7. (a) Artificial samples with labels generated by our model (the SVHN dataset). (b) Some of the actual images in the SVHN dataset.

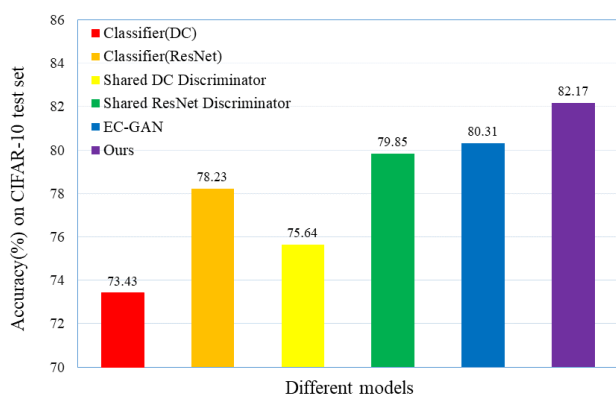


FIGURE 8. Using a 10% CIFAR-10 dataset as the model's training sample, our method's accuracy performance is compared with previous methods.

B. SVHN DATASET

The SVHN dataset [48] is obtained by collecting door numbers from Google Street View, and each sample contains a set of "0-9" Arabic numerals. The training and test sets have 73257 and 26032 images, respectively (as shown in Fig. 7(b)). During the training period, we use 5%, 10%, 15%, 20%, and 25% of the SVHN dataset as training samples to train our model to simulate real scenarios with limited datasets and then evaluate our performance on the test set.

Table 1 shows that our method improves accuracy by 1.54% compared to the best-performing EC-GAN when using 5% of the SVHN dataset as training samples. Our method outperforms other models regardless of the percentage of the SVHN dataset used as training samples. Comparing our approach with sharing ResNet discriminator and sharing DC discriminator shows that introducing an independent network for image classification on top of ACGAN is practical.

As shown in Table 2, when we train the model using different percentages of SVHN datasets, both the sample-intrinsic consistency loss (SICL) and uncertainty weighting methods proposed (UW) in this paper have about 1% improvement over the original ResNet classification network. The model performs best when the two methods are combined, demonstrating the sample's intrinsic consistency loss effectiveness with the uncertainty weighting method.

Fig 5 shows the image classification results of our model, classifier, and EC-GAN for each round after training with 10% of the SVHN dataset as the training sample. It can be seen that our model can achieve better classification results in the early stage of training. And the best classification results are performed at the end of the training compared with the classifier and EC-GAN, which shows that our model can improve the accuracy and robustness of classification.

As shown in Fig 6, we use a 5% SVHN dataset as the training sample to train our model, which aims to mimic the sample-constrained scenarios in realistic scenarios. As can be seen, our model achieves the most optimal classification accuracy compared to various previous models. Even when compared with the best-performing EC-GAN model, our model improves by 1.54 percentage points over it.

As shown in Fig 7: it is easy to see that most of the SVHN artificial images (with labels) generated have the same underlying distribution as the actual training images. It has different figures and critical features.

C. CIFAR-10 DATASET

The CIFAR-10 dataset [49] is a small dataset for recognizing generic objects and image classification, which contains ten classes of RGB color images collected by Alex Krizhevsky and Ilya Sutskever, students of Hinton. The training and test sets contain 60,000 and 10,000 photos, respectively (Figure 9(b)). During the training period, we used 10%, 15%,

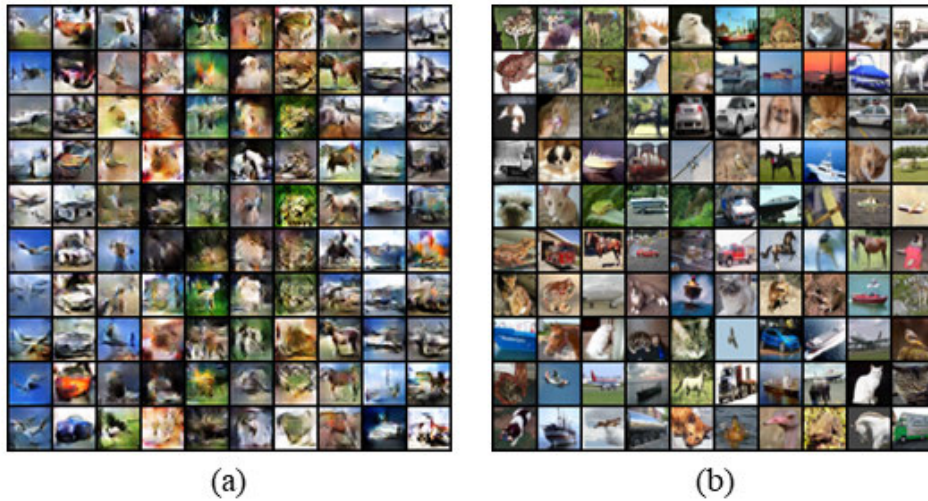


FIGURE 9. (a) Some artificial CIFAR-10 (with labels) samples generated during model training are displayed. (b) Some of the actual images in the CIFAR-10 dataset.

20%, 25%, and 30% of the CIFAR-10 dataset to train our model to simulate natural scenes with a limited dataset and evaluated our model on the test set.

Table 3 shows that our method outperforms Shared ResNet Discriminator and Shared DC Discriminator when using 10% of CIFAR-10 data as training samples, proving that our image generation model is more “perfect.” In addition, compared with the recently proposed small-sample image classification model EC-GAN, our model outperforms it by 1.86% when using 10% CIFAR-10 data as the training sample. Moreover, our model achieves better classification performance regardless of the percentage of CIFAR-10 samples used as training samples, proving our method’s effectiveness.

Fig 8 shows a schematic comparison of the accuracy of each model, given that we use 10% of the SVHN dataset as the training sample. As can be seen, our model achieves the best classification accuracy compared to various previous models. And compared with the best-performing EC-GAN model, our model also improves by 1.86 percentage points over it.

According to the results in Table 4, the sample-intrinsic consistency loss (SICL) and uncertainty weighting methods (UW) proposed in this paper can improve the performance by about 1% each when the models are trained using different proportions of SVHN datasets (relative to the original ResNet classification network). The model performs best when these two methods are combined, indicating that the sample intrinsic consistency loss and uncertainty weighting methods are effective.

We take out the CIFAR-10 data generated in the last round of the model to compare with the actual samples; as shown in Fig 9, most of the CIFAR-10 pieces with labels generated by our model are clear and have the essential characteristics of each category.

TABLE 4. The effect of the SICL and UW methods in training models using different percentages of CIFAR-10 datasets.

Method	10%	15%	20%	25%	30%
Classifier(ResNet)	78.23	81.16	86.32	88.27	89.92
Classifier(ResNet)+SICL	80.17	82.73	87.58	89.14	90.87
Classifier(ResNet)+UW	79.38	82.48	87.69	89.35	90.65
Ours	82.17	85.42	88.95	90.37	91.27

V. DISCUSSION

In the above experiments, limited training samples are simulated by adjusting the proportional size of the SVHN dataset and CIFAR-10 dataset used to participate in model training. Our model achieves 89.15% and 94.59% accuracy with 5% and 25% of the SVHN dataset as training samples, respectively, and 82.17% and 91.27% accuracy with 10% and 30% of the CIFAR-10 dataset as training samples, respectively.

These experimental results demonstrate that generating artificial images to supplement actual training samples during model training is effective. Using the sample-wise consistency loss to explore the differences between sample label information and intrinsic sample features and weighting the original supervised loss with the uncertainty weighting method is effective. However, from the experimental process and results, as the number of training samples used in the model increases, the improvement of our model’s classification performance becomes less and less noticeable. This may be due to the low quality of the generated artificial images. Therefore, our future work should consider developing more reliable artificial samples to supplement small-scale datasets in real-world tasks to improve the model’s classification ability with small-scale training samples in real scenarios.

VI. CONCLUSION

This paper proposes a new fully-supervised low-sample image classification model to improve the poor classification performance of models caused by low-sample or sample-limited datasets in real-world scenarios. This paper presents two new methods: Sample Intrinsic Consistency Loss (SICL) and Uncertainty Weighting (UW). During training, the model will generate some artificial samples to supplement the limited number of actual labeled samples and update the model parameters more effectively through the joint action of Sample Intrinsic Consistency Loss (SICL) and weighted supervised loss. The experimental results show that our method is practical and effective in improving the image classification of small-scale datasets. In addition, the proposed Sample Intrinsic Consistency Loss and Uncertainty Weighting methods can be combined with other classification models to enhance their performance.

REFERENCES

- [1] H. Ganster, P. Pinz, R. Rohrer, E. Wildling, M. Binder, and H. Kittler, "Automated melanoma recognition," *IEEE Trans. Med. Imag.*, vol. 20, no. 3, pp. 233–239, Mar. 2001.
- [2] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.
- [3] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
- [4] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
- [5] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proc. AAAI Conf. Artif. Intell.*, vol. 31, no. 1, 2017, pp. 1–7.
- [6] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.
- [7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
- [8] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos, "Image segmentation using deep learning: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 7, pp. 3523–3542, Jul. 2022.
- [9] K. Shih, C. Chiu, J. Lin, and Y. Bu, "Real-time object detection with reduced region proposal network via multi-feature concatenation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 31, no. 6, pp. 2164–2173, Jun. 2020.
- [10] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [11] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. 13th Eur. Conf. Comput. Vis. (ECCV)*, Zurich, Switzerland: Springer, Sep. 2014, pp. 818–833.
- [12] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.
- [13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [15] P. R. Jeyaraj and E. R. S. Nadar, "Medical image annotation and classification employing pyramid feature specific lightweight deep convolution neural network," *Comput. Methods Biomechanics Biomed. Eng., Imag. Visualizat.*, vol. 2023, pp. 1–12, Feb. 2023.
- [16] O. Jabari, Y. Ayalew, and T. Motshegwa, "Semi-automated X-ray transmission image annotation using data-efficient convolutional neural networks and cooperative machine learning," in *Proc. 5th Int. Conf. Video Image Process.*, Dec. 2021, pp. 205–214.
- [17] J. Cai, F. Gan, X. Cao, and W. Liu, "Signal modulation classification based on the transformer network," *IEEE Trans. Cognit. Commun. Netw.*, vol. 8, no. 3, pp. 1348–1357, Sep. 2022.
- [18] S. Hou, H. Shi, X. Cao, X. Zhang, and L. Jiao, "Hyperspectral imagery classification based on contrastive learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2021.
- [19] A. Haque, "EC-GAN: Low-sample classification using semi-supervised algorithms and GANs (student abstract)," in *Proc. AAAI Conf. Artif. Intell.*, vol. 35, no. 18, 2021, pp. 15797–15798.
- [20] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, 2020.
- [21] D. Liang, R. G. Krishnan, M. D. Hoffman, and T. Jebara, "Variational autoencoders for collaborative filtering," in *Proc. World Wide Web Conf. (WWW)*, 2018, pp. 689–698.
- [22] A.-A.-Z. Imran and D. Terzopoulos, "Multi-adversarial variational autoencoder nets for simultaneous image generation and classification," in *Deep Learning Applications*, vol. 2, 2021, pp. 249–271.
- [23] A. Vahdat and J. Kautz, "NVAE: A deep hierarchical variational autoencoder," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 33, 2020, pp. 19667–19679.
- [24] A. Sagar, "Generate high resolution images with generative variational autoencoder," 2020, *arXiv:2008.10399*.
- [25] M. Mirza and S. Osindero, "Conditional generative adversarial nets," 2014, *arXiv:1411.1784*.
- [26] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.
- [27] A. Brock, J. Donahue, and K. Simonyan, "Large scale GAN training for high fidelity natural image synthesis," 2018, *arXiv:1809.11096*.
- [28] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 2642–2651.
- [29] R. Li, L. Wang, P. N. Suganthan, and O. Sourina, "Sample-based data augmentation based on electroencephalogram intrinsic characteristics," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 10, pp. 4996–5003, Oct. 2022.
- [30] V. López, A. Fernández, S. García, V. Palade, and F. Herrera, "An insight into classification with imbalanced data: Empirical results and current trends on using data intrinsic characteristics," *Inf. Sci.*, vol. 250, pp. 113–141, Nov. 2013.
- [31] B. A. Manto and S. S. Khurana, "Static, dynamic and intrinsic features based Android malware detection using machine learning," in *Proc. ICRIC Recent Innov. Comput. Cham, Switzerland: Springer*, 2020, pp. 31–45.
- [32] Q. Liu, L. Yu, L. Luo, Q. Dou, and P. A. Heng, "Semi-supervised medical image classification with relation-driven self-ensembling model," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3429–3440, Nov. 2020.
- [33] Z. Zhou, C. Lu, W. Wang, W. Dang, and K. Gong, "Semi-supervised medical image classification based on attention and intrinsic features of samples," *Appl. Sci.*, vol. 12, no. 13, p. 6726, Jul. 2022.
- [34] P.-T. de Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Ann. Operations Res.*, vol. 134, no. 1, pp. 19–67, Feb. 2005.
- [35] J. Xu, Z. Zhang, T. Friedman, Y. Liang, and G. Broeck, "A semantic loss function for deep learning with symbolic knowledge," in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 5502–5511.
- [36] S. Maharjan, A. Alsadoon, P. W. C. Prasad, T. Al-Dalain, and O. H. Alsadoon, "A novel enhanced softmax loss function for brain tumour detection using deep learning," *J. Neurosci. Methods*, vol. 330, Jan. 2020, Art. no. 108520.
- [37] L. Chen, H. Qu, J. Zhao, B. Chen, and J. C. Principe, "Efficient and robust deep learning with coreentropy-induced loss function," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1019–1031, May 2016.
- [38] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2980–2988.
- [39] N. Sergievskiy and A. Pomarev, "Reduced focal loss: 1st place solution to xView object detection in satellite imagery," 2019, *arXiv:1903.01347*.

- [40] E. Denton, S. Gross, and R. Fergus, "Semi-supervised learning with context-conditional generative adversarial networks," 2016, *arXiv:1611.06430*.
- [41] A. Imran and D. Terzopoulos, "Multi-adversarial variational autoencoder networks," in *Proc. 18th IEEE Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2019, pp. 777–782.
- [42] C. Li, T. Xu, J. Zhu, and B. Zhang, "Triple generative adversarial nets," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 1–11.
- [43] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen, "Improved techniques for training single-image GANs," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–10.
- [44] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," 2016, *arXiv:1606.00704*.
- [45] A. Odena, "Semi-supervised learning with generative adversarial networks," 2016, *arXiv:1606.01583*.
- [46] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [47] A. Krogh and J. Hertz, "A simple weight decay can improve generalization," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 4, 1991, pp. 1–11.
- [48] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, and A. Y. Ng, "Reading digits in natural images with unsupervised feature learning," Tech. Rep., 2011.
- [49] A. Krizhevsky, V. Nair, and G. Hinton. (2020). *The CIFAR-10 Dataset (2014)*. [Online]. Available: <http://www.cs.toronto.edu/kriz/cifar.html>



XI XIAO received the B.S. degree in software engineering from the Chengdu College, University of Electronic Science and Technology of China, in 2021. She is currently pursuing the degree with the School of Computer Science, Southwest Petroleum University. Her research interests include computer vision, deep learning, and machine learning.



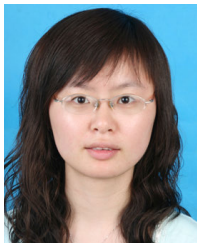
JINPENG CHEN received the B.Sc. degree in applied mathematics and statistics from Johns Hopkins University, in 2022. He is currently pursuing the Graduate degree with the Khoury College of Computer Sciences, Northeastern University. His research interests include machine learning, deep learning, and computer vision, where he aims to make significant contributions by developing novel solutions to complex problems.



ZHIGUO LI was born in Yuncheng, Shanxi, in April 1985. He received the B.Sc. degree in computer science and technology from Neijiang Normal University, Neijiang, Sichuan, China, in 2008. He is currently pursuing the Ph.D. degree with Sehan University, South Korea, in 2020. He is the Head of the Information System Department, Information Construction and Services, Neijiang Normal University. His has published articles include the Application and Expansion of Big Data of University Data Center [Information and Computer (theory version, 2020)], Visualization of Testing Data Based on ArcGIS (Computer Fans, 2018), and Issues and Solutions of University Information Construction (Science and Technology Communication, 2018). His research interests include computer networks and communications.



NAWEI ZHANG received the B.S. degree in communication engineering from Henan University, in 2020. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, China University of Petroleum, Beijing. Her research interests include computer vision, image classification, and image generation.



LINGBO LI received the B.S. degree in computer science and technology from Zhejiang Wanli College, in 2005, and the M.S. degree in software engineering from Zhejiang University, in 2007. She has been with the Information Center, Zhejiang Institute of Economics and Technology, since 2007, and has devoted herself to information construction for many years. Her research interests include information system development, computer vision, and machine learning.



SAI LI (Member, IEEE) received the Ph.D. degree in physics and electronics from the Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai, China, in 2020. He is currently with Zaozhuang University. His research interests include computer vision, machine learning, pattern recognition, hyperspectral imagery, and image processing.

...