https://doi.org/10.1007/s11704-025-50835-w

RESEARCH ARTICLE

Long-tail learning with context-aware re-sampling

Jiang-Xin SHI[1,2*], Xiao-Chao XIAO[3*], Cong-Zhong ZHU[3], Wen TAO[1,2], Wen-Yu ZHOU[4], Wei ZHU[4]✉, Yu-Feng LI[1,2]✉

1. National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
2. School of Artificial Intelligence, Nanjing University, Nanjing 210023, China
3. Suzhou Branch of China Mobile Communications Group Jiangsu Co., Ltd., Suzhou 215000, China
4. China Mobile Zijin Innovation Institute Co., Ltd., Nanjing 210000, China

Received June 16, 2025; accepted October 9, 2025

E-mail: zhuweisgs2@js.chinamobile.com; liyf@nju.edu.cn. * These authors contributed equally to this work.

Abstract

Real-world data often exhibit a long-tail class distribution, where a small subset of classes dominate the majority of the training samples, while the remaining classes suffer from severe data scarcity. Long-tail learning (LTL) aims to tackle this extreme data imbalance problem and improve the generalization across both head and tail classes. Although re-sampling offers a straightforward solution to mitigate class imbalance, prior researches have empirically shown its limited effectiveness in modern long-tail learning tasks. To overcome this limitation, we propose Context-Aware RE-sampling (CARE), a novel framework that leverages large pre-trained models to suppress irrelevant contexts as well as enrich the diversity of the training data. Specifically, CARE introduces multiple practical implementations: CARE-DS, which integrates DINO and SAM to segment and transplant objects across images, generating diverse samples while preserving semantic consistency, and CARE-DM, which utilizes diffusion models to synthesize contextually diverse samples conditioned on original images and textual prompts. Extensive experiments demonstrate that CARE effectively mitigates performance deterioration for both head and tail classes, achieving significant generalization improvements over conventional re-sampling methods.

Keywords

long-tail learning; re-sampling; class-imbalanced learning; data augmentation

1 Introduction

Recent years have witnessed well-established deep neural networks applied on various domains, particularly when it is trained upon extensive elaborated datasets [1–3]. However, despite the success, the practical application often encounters a significant challenge posed by the real-world data: the presence of a long-tail class distribution [4–6]. Such long-tail data poses two main challenges: 1) the class-imbalanced issue, causing the model to exhibit bias towards those dominated head classes; 2) the scarcity of data related to those rare tail classes, which hinders effective generalization of the model [7–10].

Re-sampling is a classic and commonly applied strategy to address the class-imbalanced problem [11,12]. By simply creating replicates of the samples of minority classes, it aims to rectify the class distribution and estimate an unbiased model. Unfortunately, previous empirical evidence suggests that re-sampling methods often limit the effects when applied to modern long-tail datasets [6,13], and are even worse than not being utilized. Previous studies subjectively attribute the limited performance of re-sampling to the risk of overfitting [6,14]. Nevertheless, the reasons behind this phenomenon remain inadequately explained in most of the existing literature. To mitigate the negative effect of re-sampling, two-stage learning is proposed, which adopts re-sampling in the later or the second stage of the whole training process. The representative two-stage learning approaches include DRS [15], cRT [6], and BBN [14]. Despite that two-stage learning can avoid the impact of re-sampling at the first stage, it still adopts re-sampling at the second stage and can inevitably fall into the overfitting problem.

To better understand the effects of re-sampling methods, one prior study [16] conducts a series of empirical investigations into the re-sampling strategies. It discovers that re-sampling does not necessarily work or fail on long-tail datasets, as it leads to absolutely contrasting effects on different long-tail datasets. As shown in Fig. 1, it helps the learning on the MNIST-LT dataset, with more balanced sampling yielding more performance improvements. In contrast, it harms the learning on CIFAR100-LT, and more balanced sampling leads to more performance deterioration. Furthermore, the additional empirical results infer that the effectiveness of re-sampling depends on the semantic relevance between training samples and their corresponding target labels [16]. Specifically, re-sampling proves

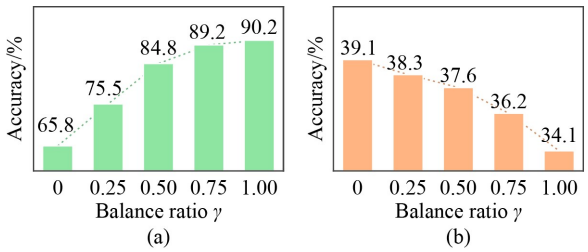


Fig. 1 The contrasting effects of re-sampling on different long-tail datasets [16], with larger balance ratio γ representing more balanced sampling strategies. (a) MNIST-LT; (b) CIFAR100-LT

advantageous when training samples exhibit high semantic relevance; otherwise, re-sampling suffers from oversampling redundant unrelated contexts, and is even worse than uniform sampling (directly sampling from the original long-tailed distributed data). It is worth mentioning that "irrelevant contexts" refer to the semantically unrelated parts of the training examples. In contrast, "related contents" refer to the semantically related parts of the examples. We give an example in Fig. 2. Interestingly, the previous works [16] point out when there exist irrelevant contexts, re-sampling yields inferior representations compared to uniform sampling, which demonstrates the negative impact of irrelevant context on re-sampling methods.

We identify that the success of re-sampling hinges on the semantic relevance between training samples and their labels. When samples exhibit high relevance to their class (e.g., clean foreground objects), re-sampling usually improves generalization; conversely, irrelevant contexts (e.g., cluttered backgrounds) lead to spurious correlations and performance degradation. Motivated by this insight, we propose Context-Aware RE-sampling (CARE), a novel framework that leverages off-the-shelf pre-trained models to generate diverse training data while suppressing irrelevant contexts (e.g., the background and the semantically irrelevant foregrounds). We first introduce CARE-DS, which combines DINO [17] and SAM [18] to segment and transplant semantically relevant objects to different backgrounds, thus enhancing diversity while preserving semantic information. Specifically, it first leverages DINO [17] to obtain a rough box coordinate of the relevant objects, and then treats the coordinate as a prompt for the SAM model to extract the semantically relevant objects more precisely. Then, it pastes the

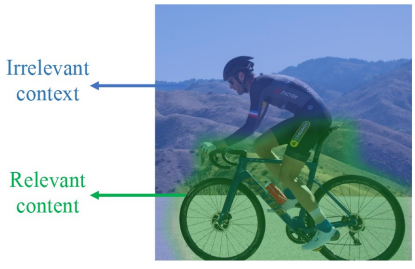


Fig. 2 Explanation of irrelevant context and relevant content. When identifying the "bicycle" in the example image, the irrelevant context refers to the semantically unrelated parts (e.g., the background and the irrelevant rider), while relevant content refers to the semantically related parts (e.g., the bicycle)

separated objects onto other randomly chosen backgrounds to generate diverse novel samples. This approach encourages the model not to focus on the irrelevant contexts and to acquire more discriminative results on both head and tail classes.

However, the performance of CARE-DS relies on the capabilities of both the DINO and the SAM models. If one of the model performs unsatisfactorily, the overall generated quality may be negatively affected. Therefore, we further propose CARE-DM, another implementation of context-aware re-sampling that uses stable diffusion models. Specifically, it synthesizes diverse samples conditioned on original images and textual prompts, such as "a photo of a cat with real-world background." Moreover, it generates more diverse samples especially for tail classes, thus re-balancing the skewed training data distributions. After generating adequate training samples, it combines the raw dataset and the generated samples and jointly trains a model for final predictions. Compared to CARE-DS, the proposed CARE-DM is more convenient for deployments by leveraging foundational generative model and conditional guidance. Although the data preparing phase requires more time cost, the final performance of CARE-DM is significantly superior to CARE-DS. Experimental results on multiple long-tail datasets demonstrate that CARE achieves competitive performance on both head and tail classes.

The key contributions of this work can be briefly summarized as follows:

- We review the efficacy of re-sampling that depends on semantic label relevance, thus highlighting the necessity of context-aware re-sampling strategies.
- We propose CARE, the first framework to explicitly suppress irrelevant contexts in re-sampling using pre-trained models. CARE includes two versions, i.e., CARE-DS and CARE-DM.
- We conduct extensive experiments to verify the effectiveness of the proposed module, which achieves an average performance gain of more than 2% in accuracy.

The rest of the paper is organized as follows. Section 2 delves into an investigation of the effects of re-sampling approaches. Section 3 presents the details of the proposed context-aware re-sampling method. Section 4 provides a brief review of the related works. Section 5 gives a conclusion of the paper.

2 Related work

2.1 Re-sampling and re-weighting

Re-sampling is widely adopted in addressing class-imbalanced problems [11,12]. Typically, re-sampling involves two main approaches: 1) Over-sampling, which replicates data from the rare classes. 2) Under-sampling, which discards a proportion of data from the frequent classes. However, conventional re-sampling methods often encounter challenges when confront with highly skewed class distributions. Previous researches propose that under-sampling may lead to the loss of valuable information, thereby inevitably degrading the model performance, while over-sampling tends to induce overfitting issues, particularly on the rare classes [14,15]. Another strategy, re-weighting, aims to generate more balanced predictions by

adjusting the losses for different classes [19,20]. An intuitive approach involves assigning weights to each training sample based on the inverse of its class frequency [4]. Similar to re-sampling, re-weighting tends to achieve better results for tail classes at the expense of the head-class performances [7].

Recent advancement [6] discovers that re-sampling is beneficial for classifier learning but has negative impacts on representation learning. Consequently, two-stage approaches [6,14,15] incorporate it in the late stage of the whole training process to mitigate its impacts on the representation. In contrast, our empirical study suggests that re-sampling can be effective, as long as there are no irrelevant contexts present. Its failures in some cases primarily stem from the unexpected overfitting towards the over-sampled, redundant contexts. When applied with context-aware augmentation, class-balanced re-sampling can achieve competitive performance on long-tail datasets.

2.2 Data augmentation

Various data augmentation approaches have been proposed to enhance model generalization capabilities. In research fields such as contrastive learning [21,22], curriculum learning [23], meta-learning [24], semi-supervised learning [25,26], and noisy-label learning [27–29], data augmentation strategies have demonstrated the effectiveness in improving the generalization of tail classes. MiSLAS [30] investigates the mixup technique [31] in long-tail learning and observes a positive effect of mixup on representation learning. However, it brings about a negative or negligible effect on classifier learning. Remix [32] adapts the mixup method to a re-balanced version by introducing a disproportionately higher weight for the tail class, thereby assigning the mixed label in favor of the tail class,

Head-to-tail knowledge transfer strategies have also been proposed for data augmentation in long-tail learning. Major-to-minor translation (M2m) [33] employs over-sampling and translates the head-class samples to replace the duplicated tail-class samples through adversarial perturbations. CAM-BS [13] separates the foreground and background of each sample, augmenting the foreground part through flipping, translating, rotating, or scaling. However, it overlooks the limited generalization ability of the learned model on tail classes, which may yield incredible separations. CMO [34] applies CutMix [35] by cutting out random regions of a head-class sample and filling these regions with another sample from tail classes. By this means, it enriches the contexts of the tail data. Nevertheless, the random cutout operation does not necessarily distinguish the semantically related objects or irrelevant contexts. In contrast, our method utilizes the off-the-shelf pre-trained model to separate related objects from contexts in training samples. It then pastes the well-separated objects onto more backgrounds to generate diverse novel samples.

2.3 Pretrained models

Recent advances in self-supervised learning and generative models present novel approaches for data augmentation. DINO [36] leverages vision transformers to extract semantically meaningful object representations without manual annotations. The ability of DINO on capturing object-level features has been exploited in segmentation and augmentation tasks [17,36]. Similarly, the Segment Anything Model (SAM) [18] provides a powerful and promptable segmentation framework, which is capable of precisely segmenting objects in complex scenarios. Recent works have utilized SAM for data augmentation by extracting and recombining salient regions, thus enhancing sample diversity while preserving semantic consistency [37,38].

Diffusion models [39–41] have emerged as a powerful foundation model to synthesize high-quality and diverse samples. Unlike traditional augmentation techniques, diffusion models can synthesize contextually coherent images conditioned on textual or visual prompts. For instance, Guided Imagination Framework (GIF) [42] shows that diffusion-generated samples can improve generalization of small-scale datasets by preserving class-relevant features. Moreover, recent studies [43,44] demonstrate the effectiveness of diffusion models in long-tail learning by generating tail-class samples with different stategies. However, the effectiveness of diffusion models on avoiding spurious correlations has not been studied. In this work, we studies how to leverage well-trained models such as DINO, SAM, and stable-diffusion models for mitigating irrelevant contexts and generating more diverse samples.

3 Context-aware re-sampling

3.1 Preliminaries

Given a training dataset $\mathcal{D} = \{\boldsymbol{x}_i, y_i\}_{i=1}^N$, where $\boldsymbol{x}_i$ is a training sample and $y_i \in C = [K] = \{1, \ldots, K\}$ is the corresponding label. We assume that the training data follow a long-tail class distribution, i.e., the class prior distribution $\mathbb{P}(y)$ is long-tailed. In this case, a majority of classes have a very low probability of occurrence. Moreover, we define the imbalance ratio as $\rho = \max_y \mathbb{P}(y) / \min_y \mathbb{P}(y)$ to indicate the long-tailedness of data. Classes with high $\mathbb{P}(y)$ are referred to as head classes, while others are referred to as tail classes.

In conventional settings, where the training and test data obey a consistent distribution, Empirical Risk Minimization (ERM) is widely used upon the training data to achieve an empirical estimate of the underlying test data distribution. A typical approach is to minimize the Cross-Entropy (CE) loss for each training sample as following:

$$\mathcal{L}_{\mathrm{CE}} = \frac{1}{N} \sum_{i=1}^{N} -\log \frac{\exp(z_{y_i})}{\sum_{k=1}^{K} \exp(z_k)}, \tag{1}$$

where z_k denotes the predictive logit of sample $\boldsymbol{x}$ on class k. However, this prevalent approach neglects the issue of class imbalance, leading to learned models biased toward head classes [8,45].

To deal with the class-imbalance problem, the re-sampling strategy assigns a probability of being selected for each training sample according to its class frequency [6]. The probability of sampling a data point $\boldsymbol{x}_i$ can be written as:

$$p(\boldsymbol{x}_i) = \frac{n_{y_i}^{-\gamma}}{\sum_{j=1}^{N} n_{y_j}^{-\gamma}}, \tag{2}$$

where n_k denotes the frequency of class k. When $\gamma = 0$, Eq. (2)

denotes uniform sampling, where all training data share the same sampling probability. In this case, tail-class data have lower probabilities of being sampled. When $\gamma = 1$, Eq. (2) denotes class-balanced re-sampling, where the sampling probability is related to the reciprocal of the corresponding class frequency, and each class has an equal probability of occurence.

By studying the effects of re-sampling methods in different scenarios [16], we can draw a conclusion: when the training samples contain irrelevant contexts, simply over-sampling the tail-class samples might cause the model to unexpectedly focus on these redundant contexts, thereby resulting in the overfitting problem. Therefore, we naturally raise a question: can we separate the irrelevant contexts to avoid the overfitting problem? Inspired by this motivation, we design a context-aware re-sampling module by utilizing off-the-shelf pre-trained segmentation models to separate the contexts as well as enrich the diversity of the training data. Note that with the rapid development of large pre-trained models [46,47], how to use such off-the-shelf models to assist in the training of small models in specific fields has also attracted widespread concerns. However, they mainly focus on model distillation or compression [48], which often requires extensive training epochs.

3.2 Re-sampling with enriching the contexts

The segment-anything (SAM) model [18] has been proposed recently, which can help separate the objects or backgrounds for a given image with a simple prompt. Such prompts can be a set of points, a box, some masks, or a text sentence. Note that though SAM allows text prompts as inputs, it has not open-sourced the detailed implementation. Nevertheless, existing works mainly combine SAM with other text-driven recognition models, such as DINO [17]. Specifically, one can first use DINO to achieve a rough box coordinate by giving a text description of the object, and then utilize the box as a prompt for SAM to extract the detailed object masks.

To mitigate the negative effect of irrelevant contexts and avoid overfitting issues, we propose to separate the contexts of the training samples. However, the absence of the contexts gives rise to the change in the style of the images, for example, a pure black background. To avoid this potential issue, we utilize the extracted contexts to enrich other samples from the same class. In this way, it keeps the semantic relevant objects unchanged while enriching the diversity of the irrelevant contexts, thereby preventing the model from overfitting to the contexts.

Formally, given an image $\boldsymbol{x}_i$, we calculate its mask $\boldsymbol{M}_i$ that is semantically related to the corresponding label. Here $\boldsymbol{M}_i$ is a matrix of the same size as $\boldsymbol{x}_i$ with boolean values. The positive value indicates that the corresponding pixel belongs to the semantically related object, while the other pixels belong to the contexts.

After calculating the mask, we obtain the object by $\boldsymbol{M}_i \odot \boldsymbol{x}_i$. Moreover, we replace the contexts with a background image of the other training samples, i.e., $(1-\boldsymbol{M}_i)\odot\boldsymbol{x}_j$. In this way, we generate more diverse novel samples by keeping the semantically related objects and simulating the objects within various contexts. To further encourage diversity, we apply a random fusing factor λ for generating novel images. Formally, we have

$$\lambda \sim \text{Uniform}(0,1). \tag{3}$$

$$\tilde{\boldsymbol{x}}_i = \boldsymbol{M}_i \odot \boldsymbol{x}_i + (1-\boldsymbol{M}_i)\odot(\lambda \boldsymbol{x}_j + (1-\lambda)\boldsymbol{x}_i). \tag{4}$$

Different from previous mixup-based methods [31,35], our method does not change the target label, because the pasted background is not related to the semantics of any class labels. Finally, we replace x_i with $\tilde{\boldsymbol{x}}_i$ to calculate the corresponding training loss. We call this framework Context-Aware Re-sampling with DINO and SAM (CARE-DS). Figure 3 gives a brief overview of the proposed module CARE-DS.

Note that the context-aware augmentation is only applied in the training stage. While in the inference phase, we predict from the original image to ensure the inference efficiency.

3.3 Re-sampling with diverse image generation

While CARE-DS mitigates irrelevant contexts by physically segmenting and transplanting objects, we further propose CARE-DM, a generative approach that leverages diffusion models to synthesize diverse and high-quality samples for tail classes. Unlike traditional augmentation or pasting-based strategies, diffusion models inherently generate contextually coherent images while preserving semantic relevance to the corresponding class labels.

Given an image with label y_i, we employ a pre-trained text-conditioned diffusion model to synthesize new samples. The text prompt is structured as "a photo of a {class} with real-world background", where {class} denotes the textual descriptor of y_i (e.g., "cat" or "dog"). This prompt ensures the generated images retain class-discriminative features while varying backgrounds naturally.

Moreover, in order to ensure the generation quantity as well as control the augmentation scale, we generate more samples for tail-class samples by reducing the imbalance ratio ($\rho = n_{\max}/n_{\min}$) to $\sqrt{\rho}$. For instance, if the original long-tail dataset contains several classes with number of samples per class ranging from 100 to 1, then we generate enough samples to adjust the class frequency to ranging from 100 to 10. We call this framework Context-Aware Re-sampling with Diffusion Model (CARE-DM), which is illustrated in Fig. 4. Compared with CARE-DS which relies on object-background composition and may occasionally produce unnatural boundaries,

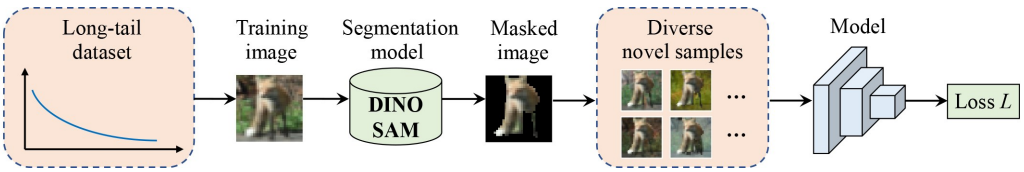


Fig. 3 An overview of context-aware re-sampling with DINO and SAM (CARE-DS)

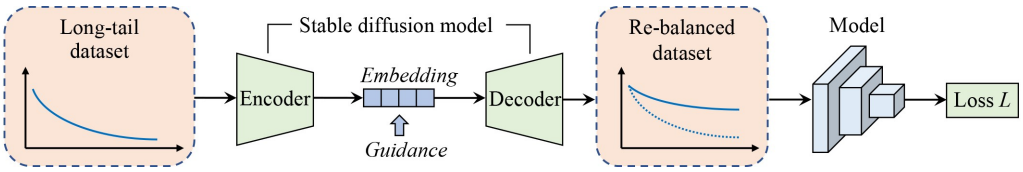


Fig. 4 An overview of context-aware re-sampling with diffusion model (CARE-DM)

using Diffusion models can generate globally coherent images and lead to more satisfactory data augmentation effects.

3.4 Final objective

The final learning objective is presented in Eqs. (5)−(7).

$$(\boldsymbol{x}, y) \sim \hat{\mathcal{D}}, \tag{5}$$

$$z_k = \Phi(\boldsymbol{x}), \tag{6}$$

$$\mathcal{L}_{CE} = \frac{1}{\hat{N}} \sum_{i=1}^{\hat{N}} -\log \frac{\exp(z_{y_i})}{\sum_{k=1}^{K} \exp(z_k)}, \tag{7}$$

where $\hat{\mathcal{D}}$ refers to the augmented dataset, which contains the original training samples and the generated novel samples. z_k denotes the predictive logit of sample $\boldsymbol{x}$ on class k using the learned model Φ. $\hat{N}$ denotes the number of training samples in the augmented dataset. We optimize the cross-entropy loss considering that the class distribution of the training dataset is rectified to a more balanced distribution.

4 Empirical results

4.1 Experimental settings

We demonstrate the effectiveness of the proposed method CARE by comparing it with different kinds of long-tail learning methods, including:

- Re-sampling or re-weighting methods, such as Focal Loss [49], CB-Focal [7], CE-DRS [15], CE-DRW [15], LDAM-DRW [15], cRT [6], LWS [6], and BBN [14],
- Data augmentation methods, such as Mixup [31], Remix [32], M2m [33], CAM-BS [13], CMO [34], and CSA [16].

We conduct experiments on CIFAR10-LT [15], CIFAR100-LT [15], the long-tail versions of CIFAR datasets by sampling from the raw dataset with an imbalance ratio ρ. Following previous works [14,15], we conduct experiments with $\rho \in \{100, 50, 10\}$. For CIFAR-100, we split the classes into three shots: head (0−35 classes), medium (36−70 classes), and tail (71−99 classes) classes. For CIFAR-10, the classes are also split into three shots: head (0−2), medium (3−6), and tail (7−9) classes. For each dataset, we evaluate the model on another corresponding class-balanced test set by calculating the overall prediction accuracy.

We use ResNet-32 as the backbone network and train it using standard SGD with a momentum of 0.9, a weight decay of 2×10^{-4}, and a batch size of 128. The model is trained for 200 epochs. The initial learning rate is set to 0.2 and is annealed by a factor of 10 at the 160th and the 180th epochs. We train each model with 1 NVIDIA GeForce RTX 3090. In all experiments, we utilize the context-aware augmentation for the whole of 200 epochs. Following deferred re-sampling (DRS), we adopt re-sampling at the 160th epoch to obtain a balanced classifier.

4.2 Results on long-tail datasets

The results for CIFAR10-LT and CIFAR100-LT are summarized in Table 1. We report the results under imbalance ratio $\rho = 100, 50, 10$. As shown in the results, CARE-DS outperforms most of the existing methods, especially under high imbalance ratios. CARE-DM even achieves the highest performance compared with the baseline methods under all scenarios.

Specifically, the methods based on re-sampling or re-weighting such as DRS, DRW, cRT, and LWS can ease the class-imbalanced problem to some degree but the performance gain is limited due to the negative effects of re-sampling. The methods based on data augmentations, such as CAM-BS and CMO, achieve higher performance. CAM-BS adopts the class activation map (CAM) to separate the objects and the contexts of each sample and then applies balanced re-sampling. Nevertheless, it neglects that the learned

Table 1 Test accuracy (%) on CIFAR datasets with various imbalanced ratios (ρ=100, 50, 10)

Method	CIFAR100-LT			CIFAR10-LT		
	100	50	10	100	50	10
CE	38.3	43.9	55.7	70.4	74.8	86.4
Focal Loss [49]	38.4	44.3	55.8	70.4	76.7	86.7
CB-Focal [7]	39.6	45.2	58.0	74.6	79.3	87.1
CE-DRS [15]	41.6	45.5	58.1	75.6	79.8	87.4
CE-DRW [15]	41.5	45.3	58.1	76.3	80.0	87.6
LDAM-DRW [15]	42.0	46.6	58.7	77.0	81.0	88.2
cRT [6]	42.3	46.8	58.1	75.7	80.4	88.3
LWS [6]	42.3	46.4	58.1	73.0	78.5	87.7
BBN [14]	42.6	47.0	59.1	79.8	82.2	88.3
mixup [31]	39.5	45.0	58.0	73.1	77.8	87.1
Remix [32]	41.9	–	59.4	75.4	–	88.2
M2m [33]	43.5	–	57.6	79.1	–	87.5
CAM-BS [13]	41.7	46.0	–	75.4	81.4	–
CMO [34]	43.9	48.3	59.5	–	–	–
CSA [16]	45.8	49.6	61.3	80.6	84.3	90.8
CARE-DS (Ours)	46.2	51.1	61.2	79.4	83.7	89.0
CARE-DM (Ours)	**49.7**	**52.5**	**63.7**	**82.0**	**84.7**	**93.6**

model has limited generalization ability, and the separated contexts are incredible. CMO applies CutMix [35] and combines uniform sampling and balanced re-sampling. However, the simple cutout operation can not separate the objects and contexts. In contrast, our proposed method CARE-DS utilizes the off-the-shelf model to extract contexts and can consistently achieve satisfactory performance. Our method CARE-DM leverages the advantages of diffusion models and achieves superior performance in all experimental settings.

4.3 Results of different shots of classes

Apart from the main results, we also report the results of different shots of classes. The results are presented in Tables 2−7. Compared with CARE-DS, CARE-DM performs better in most of the cases. Moreover, it improves the tail-class performance particularly under high class imbalance scenarios, indicating the effectiveness of context-aware re-sampling under extreme data scarcity.

4.4 Comparison of different rectified imbalance ratios

To improve the diversity of the generated samples while ensuring the efficiency of generation, we propose to adjust the imbalance ratio ($\rho = n_{\max}/n_{\min}$) to $\sqrt{\rho}$. We conduct a comparison experiment of different rectified imbalance ratios ($\sqrt{\rho}$ vs. 1) and report the results in Table 8. The results show that adjusting the imbalance ratio to 1 does not lead to further performance improvement. This indicates that adjusting the imbalance ratio from ρ to $\sqrt{\rho}$ is sufficient for enhancing performance, while generating more samples has limited benefits. We add the comparison results in the updated manuscript.

Table 2 Test accuracy (%) of different shots of classes on CIFAR-100-IR100 dataset

	Overall	Head	Medium	Tail
CARE-DS	46.2	63.6	48.3	22.2
CARE-DM	49.7 (+3.5)	65.5 (+1.9)	51.2 (+2.9)	26.1 (+3.9)

Table 3 Test accuracy (%) of different shots of classes on CIFAR-100-IR50 dataset

	Overall	Head	Medium	Tail
CARE-DS	51.1	64.2	53.3	32.0
CARE-DM	52.5 (+1.4)	66.1 (+1.9)	54.3 (+1.0)	33.1 (+1.1)

Table 4 Test accuracy (%) of different shots of classes on CIFAR-100-IR10 dataset

	Overall	Head	Medium	Tail
CARE-DS	61.2	67.6	62.1	52.1
CARE-DM	63.7 (+2.5)	66.0 (−1.6)	64.8 (+2.7)	54.2 (+2.1)

Table 5 Test accuracy (%) of different shots of classes on CIFAR-10-IR100 dataset

	Overall	Head	Medium	Tail
CARE-DS	79.4	91.7	78.1	68.9
CARE-DM	82.0 (+2.6)	91.4 (−0.3)	80.9 (+2.8)	70.6 (+1.7)

Table 6 Test accuracy (%) of different shots of classes on CIFAR-10-IR50 dataset

	Overall	Head	Medium	Tail
CARE-DS	83.7	91.2	81.8	78.7
CARE-DM	84.7 (+1.0)	92.6 (+1.4)	83.4 (+1.6)	82.6 (+3.9)

Table 7 Test accuracy (%) of different shots of classes on CIFAR-10-IR10 dataset

	Overall	Head	Medium	Tail
CARE-DS	89.0	92.5	86.5	88.7
CARE-DM	90.7 (+1.7)	93.6 (+1.1)	86.9 (+0.4)	90.5 (+1.8)

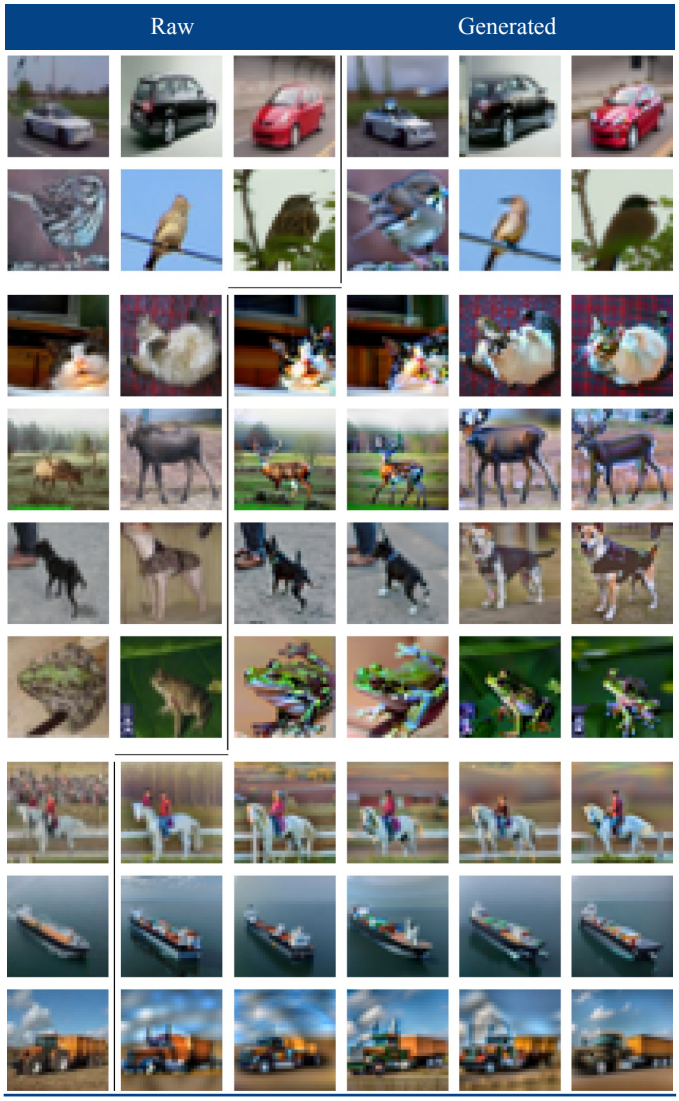
Table 8 Test accuracy (%) on CIFAR datasets with various rectified imbalanced ratios

Method	CIFAR100-LT			CIFAR10-LT		
	100	50	10	100	50	10
$\sqrt{\rho}$ (square root)	49.7	**52.5**	**63.7**	**82.0**	**84.7**	**93.6**
1 (balanced)	**49.8**	52.3	63.5	81.2	84.5	**93.6**

4.5 Visualization of generated images

In CARE-DM, we utilize the diffusion model for generating diverse and high-quality images. In Table 9, we visualize the generated images from the CIFAR10-IR100 dataset. We choose all of the augmented classes ("automobile", "bird", "cat", "deer", "dog", "frog", "horse", "ship", and "truck") in this dataset and randomly select raw images and the corresponding generated images. Note that the class "airplane" is not included, since it is the most frequent class in CIFAR10-IR100, thus CARE-DM does not require to generate new samples for this class. The visualization results in Table 9 show

Table 9 Visualization of images generated by CARE-DM on CIFAR10-IR100



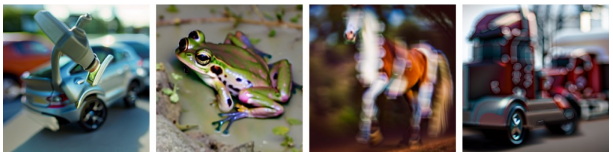


Fig. 5 Illustration of misrepresented generated samples

that the diffusion model can effectively generates diverse and high-quality images for different training images. When taking a close look at the objects in the images, it can be found that the backgrounds and irrelevant contexts are generated into multiple diverse versions. Such generating results can effectively help mitigate the overfitting to irrelevant contexts, and thereby improving the generalization ability of the learned model.

Moreover, one may be interested in the negative cases where generated samples misrepresent the sample characteristics. We have checked the generated samples and find that although most of the samples are normal, there are also misrepresented samples on both head and tail classes. We visualize the misrepresented samples in Fig. 5. For example, it may generate images of automobile with strange structure, frog with three eyes, and some blurry images of horse or truck. This issue is mainly related to the ability of the generated model. Nonetheless, the overall quality of generated samples is positive, which can also be verified by the improved model performance. We will incorporate more versatile generative model in future work.

5 Conclusion

This work rethinks the intrinsic relationship between re-sampling and its impact on long-tail learning. Motivated by the effects of re-sampling on different long-tail tasks and the sensitivity of re-sampling on irrelevant contexts, we proposed Context-Aware RE-sampling (CARE) to address the potential challenge. CARE presents a novel framework that mitigates the negative impact of irrelevant contexts through diverse data augmentation. By leveraging off-the-shelf segmentation or generative models, CARE is able to effectively separates semantically relevant objects from trivial backgrounds, and generates diverse and high-quality training samples through context-aware composition. By conducting experiments on long-tail datasets and comparing the performance with re-sampling, re-weighting, and data augmentation methods, we demonstrated the effectiveness of our proposed method CARE by a considerable margin.

Code availability statement

The source code for this work is available at the website of github.com/shijxcs/CARE.

Acknowledgements

This research was supported by the Jiangsu Science Foundation (BK20243012), the National Natural Science Foundation of China (Grant Nos. 62576162, 62576174), and Nanjing University–China Mobile Communications Group Co.,Ltd. Joint Institute.

Competing interests

The authors declare that they have no competing interests or financial conflicts to disclose.

References

[1] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521(7553): 436−444

[2] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016, 770−778

[3] Deng J, Dong W, Socher R, Li L J, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. 2009, 248−255

[4] Wang Y X, Ramanan D, Hebert M. Learning to model the tail. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017, 7032−7042

[5] Liu Z, Miao Z, Zhan X, Wang J, Gong B, Yu S X. Large-scale long-tailed recognition in an open world. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, 2532−2541

[6] Kang B, Xie S, Rohrbach M, Yan Z, Gordo A, Feng J, Kalantidis Y. Decoupling representation and classifier for long-tailed recognition. In: Proceedings of the 8th International Conference on Learning Representations. 2020

[7] Cui Y, Jia M, Lin T Y, Song Y, Belongie S. Class-balanced loss based on effective number of samples. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019, 9260−9269

[8] Ren J, Yu C, Sheng S, Ma X, Zhao H, Yi S, Li H. Balanced meta-softmax for long-tailed visual recognition. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020, 351

[9] Samuel D, Chechik G. Distributional robustness loss for long-tail learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021, 9475−9484

[10] Yang Y, Xu Z. Rethinking the value of labels for improving class-imbalanced learning. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020, 1618

[11] Chawla N V, Bowyer K W, Hall L O, Kegelmeyer W P. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 2002, 16: 321−357

[12] Liu X Y, Wu J, Zhou Z H. Exploratory undersampling for class-imbalance learning. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 2009, 39(2): 539−550

[13] Zhang Y, Wei X S, Zhou B, Wu J. Bag of tricks for long-tailed visual recognition with deep convolutional neural networks. In: Proceedings of the 35th AAAI Conference on Artificial Intelligence. 2021, 3447−3455

[14] Zhou B, Cui Q, Wei X S, Chen Z M. BBN: bilateral-branch network with cumulative learning for long-tailed visual recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 9716−9725

[15] Cao K, Wei C, Gaidon A, Arechiga N, Ma T. Learning imbalanced datasets with label-distribution-aware margin loss. In: Proceedings of the 33rd International Conference on Neural Information Processing Systems. 2019, 140

[16] Shi J X, Wei T, Xiang Y, Li Y F. How re-sampling helps for long-tail learning? In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 3307

[17] Liu S, Zeng Z, Ren T, Li F, Zhang H, Yang J, Jiang Q, Li C, Yang J, Su H, Zhu J, Zhang L. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection. In: Proceedings of the 18th European Conference on Computer Vision. 2023, 38−55

[18] Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, Xiao T, Whitehead S, Berg A C, Lo W Y, Dollár P, Girshick R. Segment anything. In: Proceedings of 2023 IEEE/CVF International Conference on Computer Vision (ICCV). 2023, 3992−4003

[19] Elkan C. The foundations of cost-sensitive learning. In: Proceedings of the 17th International Joint Conference on Artificial Intelligence. 2001, 973−978

[20] Zhou Z H, Liu X Y. Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(1): 63−77

[21] Zhou Z, Yao J, Wang Y F, Han B, Zhang Y. Contrastive learning with boosted memorization. In: Proceedings of the 39th International Conference on Machine Learning. 2022, 27367−27377

[22] Zhu J, Wang Z, Chen J, Chen Y P P, Jiang Y G. Balanced contrastive learning for long-tailed visual recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, 6898−6907

[23] Ahn S, Ko J, Yun S Y. CUDA: curriculum of data augmentation for long-tailed recognition. In: Proceedings of the 11th International Conference on Learning Representations. 2023

[24] Li S, Gong K, Liu C H, Wang Y, Qiao F, Cheng X. MetaSAug: meta semantic augmentation for long-tailed visual recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, 5208−5217

[25] Li Y F, Liang D M. Safe semi-supervised learning: a brief introduction. Frontiers of Computer Science, 2019, 13(4): 669−676

[26] Guo L Z, Jia L H, Shao J J, Li Y F. Robust semi-supervised learning in open environments. Frontiers of Computer Science, 2025, 19(8): 198345

[27] Wei T, Shi J X, Zhang M L, Li Y F. Robust long-tailed learning under label noise. Frontiers of Computer Science, 2026, 20(1): 2001321

[28] Zhou Z, Jin Y X, Li Y F. Rts: learning robustly from time series data with noisy label. Frontiers of Computer Science, 2024, 18(6): 186332

[29] Li S Y, Zhao S J, Cao Z T, Huang S J, Chen S. Robust domain adaptation with noisy and shifted label distribution. Frontiers of Computer Science, 2025, 19(3): 193310

[30] Zhong Z, Cui J, Liu S, Jia J. Improving calibration for long-tailed recognition. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021, 16484−16493

[31] Zhang H, Cissé M, Dauphin Y N, Lopez-Paz D. mixup: beyond empirical risk minimization. In: Proceedings of the 6th International

Conference on Learning Representations. 2018
[32] Chou H P, Chang S C, Pan J Y, Wei W, Juan D C. Remix: rebalanced mixup. In: Proceedings of the 16th European Conference on Computer Vision. 2020, 95−110
[33] Kim J, Jeong J, Shin J. M2m: imbalanced classification via major-to-minor translation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 13893−13902
[34] Park S, Hong Y, Heo B, Yun S, Choi J Y. The majority can help the minority: context-rich minority oversampling for long-tailed classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, 6877−6886
[35] Yun S, Han D, Chun S, Oh S J, Yoo Y, Choe J. CutMix: regularization strategy to train strong classifiers with localizable features. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019, 6022−6031
[36] Zhang H, Li F, Liu S, Zhang L, Su H, Zhu J, Ni L, Shum H Y. DINO: DETR with improved DeNoising anchor boxes for end-to-end object detection. In: Proceedings of the 11th International Conference on Learning Representations. 2023
[37] Zhang Y, Zhou T, Wang S, Liang P, Zhang Y, Chen D Z. Input augmentation with SAM: boosting medical image segmentation with segmentation foundation model. In: Proceedings of the International Conference on Medical Image Computing and Computer Assisted Intervention. 2023, 129−139
[38] Dai H, Ma C, Yan Z, Liu Z, Shi E, Li Y, Shu P, Wei X, Zhao L, Wu Z, Zeng F, Zhu D, Liu W, Li Q, Sun L, Liu S Z T, Li X. SAMAug: point prompt augmentation for segment anything model. 2023, arXiv preprint arXiv: 2307.01187
[39] Ho J, Jain A, Abbeel P. Denoising diffusion probabilistic models. In: Proceedings of the 34th International Conference on Neural Information Processing Systems. 2020, 574
[40] Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022, 10674−10685
[41] Croitoru F A, Hondru V, Ionescu R T, Shah M. Diffusion models in vision: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(9): 10850−10869
[42] Zhang Y, Zhou D, Hooi B, Wang K, Feng J. Expanding small-scale datasets with guided imagination. In: Proceedings of the 37th International Conference on Neural Information Processing Systems. 2023, 3346
[43] Shao J, Zhu K, Zhang H, Wu J. DiffuLT: diffusion for long-tail recognition without external knowledge. In: Proceedings of the 38th International Conference on Neural Information Processing Systems. 2024, 3909
[44] Zhang T, Zheng H, Yao J, Wang X, Zhou M, Zhang Y, Wang Y. Long-tailed diffusion models with oriented calibration. In: Proceedings of the 12th International Conference on Learning Representations. 2024
[45] Menon A K, Jayasumana S, Rawat A S, Jain H, Veit A, Kumar S. Long-tail learning via logit adjustment. In: Proceedings of the 9th International Conference on Learning Representations. 2021
[46] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł, Polosukhin I. Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017, 6000−6010
[47] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Houlsby N. An image is worth 16x16 words: transformers for image recognition at scale. In: Proceedings of the 9th International Conference on Learning Representations. 2021
[48] Touvron H, Cord M, Douze M, Massa F, Sablayrolles A, Jegou H. Training data-efficient image transformers & distillation through attention. In: Proceedings of the 38th International Conference on Machine Learning. 2021, 10347−10357
[49] Lin T Y, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. In: Proceedings of the IEEE International Conference on Computer Vision. 2017, 2999−3007

Jiang-Xin SHI received his BSc degree in 2020. He is currently working toward the PhD degree at the School of Artificial Intelligence, the National Key Laboratory for Novel Software Technology at Nanjing University, China. His research interests focus on long-tail learning. He served as the Program Committee Member for top-tier conferences, e.g., ICML/NeurIPS/ICLR/KDD/AAAI, and the Senior Program Committee Member for IJCAI 2025.

Xiao-Chao XIAO is a Senior Engineer and Deputy Manager of the Network Department at Suzhou Branch of Jiangsu Co., Ltd., China Mobile Communications Group. His research directions include machine learning, communication networks, etc.

Cong-Zhong ZHU obtained his Master's degree from Southeast University, China. He is currently a wireless network optimization engineer with the Suzhou Branch of China Mobile Communications Group Jiangsu Co., Ltd. His primary research focuses on electromagnetic field and related fields.

Wen TAO received his BSc degree in 2023. He is currently working toward the MSc degree at the School of Artificial Intelligence, the National Key Laboratory for Novel Software Technology at Nanjing University, China. His research interest is machine learning.

Wen-Yu ZHOU received his ME degree in Engineering from Nanjing University of Posts and Telecommunications, China. He serves as a research and development engineer in Wireless Cloud–Network R&D Department at China Mobile Zijin Innovation Institute Co., Ltd. His main research directions include deep learning and knowledge graph, etc.

Wei ZHU received his ME degree in Engineering from Nanjing University of Posts and Telecommunications, China. He is currently the Manager of the Wireless Cloud–Network R&D Department at China Mobile Zijin Innovation Institute Co., Ltd. He also serves as a member of the

Digital Twin and System Simulation Professional Committee of the China Institute of Communications and as a provincial-level expert for China Mobile. His research interests span artificial intelligence, digital-twin modeling, etc.

Yu-Feng LI received his BSc and PhD degrees in computer science from Nanjing University, China in 2006 and 2013, respectively. He is currently a full professor at the School of Artificial Intelligence, the National Key Laboratory for Novel Software Technology at Nanjing University, China. He is a member of the LAMDA group, led by Prof. Zhi-Hua Zhou (IEEE/ACM/AAAI Fellow). His research interests focus on robust and reliable machine learning. He has published over 90 academic papers in top-tier journals and conferences in the field, with around 7000 citations. He serves as journal associate/action editor for Artificial Intelligence, Machine Learning, Neural Networks, etc. He served as program co-chair for IEEE Big Comp 2020/CCML 2021, and area chairs for top-tier conferences, e.g., ICML/NeurIPS/ICLR/IJCAI. The research work has been selected for the IJCAI 2021 Early-Career Spotlight Talk. He won the PAKDD Early-Career Research Award 2024. He is the associated program co-chair for IJCAI 2025.