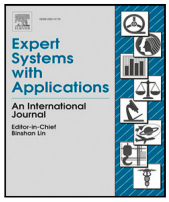
Expert Systems With Applications 246 (2024) 123138

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa



ECS-SC: Long-tailed classification via data augmentation based on easily confused sample selection and combination

Wenwei He, Junyan Xu, Jie Shi, Hong Zhao *

School of Computer Science, Minnan Normal University, Zhangzhou, Fujian, 363000, China
The Key Laboratory of Data Science and Intelligence Application, Minnan Normal University, Zhangzhou, Fujian, 363000, China

ARTICLE INFO

Keywords:
Long-tailed classification
Data augmentation
Multi-granularity
Easily confused sample

ABSTRACT

The long-tailed distribution data poses many challenges for machine learning because the tail classes are extremely scarce. Long-tailed data augmentation is a powerful technique for enriching the tail class diversity. However, existing methods often treat each class independently, assuming that classes are isolated from each other. These approaches overlook the presence of easily confused tail classes, making it challenging for models to distinguish between them accurately. In this paper, we propose a long-tailed classification method based on data augmentation, which utilizes multi-granularity knowledge to select and combine easily confused tail samples, thereby enhancing the classification performance of these samples. First, we utilize multi-granularity knowledge and semantic relation trees to build a class relation matrix. This matrix records the relationship between classes and helps the model search for easily confused classes from bilateral branch samplers. Second, we crop and combine the easily confused head and tail class samples in a foreground–background manner to generate new samples, thereby augmenting the model training. The extensive head class knowledge is transferred to the scarce tail class samples through the combination of fore-background, and the discriminative and generalized abilities of the model are improved. The experimental results affirm the effectiveness of our proposed method.

1. Introduction

Long-tailed classification, which involves imbalanced data distributions and limited samples in tail classes, is one of the research hotspots in deep learning (Patel, Jana, & Mahanta, 2023). Specifically, the long-tailed distribution means that the majority of samples are clustered with a few classes, and most classes contain only a limited number of samples, which are named head and tail classes, respectively. The long-tailed distribution is prevalent in numerous applications, such as face recognition (Krizhevsky, Sutskever, & Hinton, 2017; Wang, Liu, et al., 2023), biomedicine (Chen et al., 2023), endangered species recognition (Varghese, Jawahar, & Prince, 2023), and disaster prediction (Xue, Fei, & Ding, 2024).

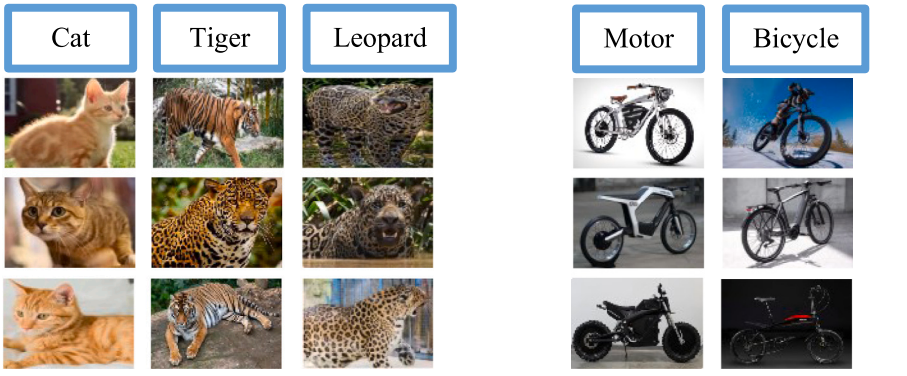
Model performance deteriorates when trained directly on long-tail datasets (He, Zhang, Ren, & Sun, 2016). The recognition accuracy and generalization ability of the tail classes are affected because these models are easily biased towards the head classes (Li & Shen, 2023). Tail classes are overfitted since their sample number cannot describe their intra-class diversity (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). Therefore, addressing long-tailed distribution in deep learning models poses an urgent challenge (Wang, Zhou, Li, & an Zhang, 2023).

Numerous studies have been presented to mitigate the issue of long-tailed distribution, including re-balancing and data augmentation methods. Re-balancing methods include re-weighting and re-sampling techniques (Deng, Zhong, Ren, Zeng, & Zhang, 2016; Lin, Goyal, Girshick, He, & Dollár, 2017). Re-weighting techniques regulate the loss function weight of tail classes by imposing higher penalties to make the model focus on the tail classes (Hong et al., 2021; Park, Lim, Jeon, & Choi, 2021). Unlike re-weighting methods that adjust the weight of loss function, the re-sampling techniques promote training fairness by repeatedly sampling sample images, including under-sampling and over-sampling strategies. The over-fitting of the tail class is a key issue due to their limited intra-class diversity, although re-balancing methods alleviate the skewness of the model.

Data augmentation in long-tailed classification alleviates effectively the issue, divided into inter-class data augmentation methods (Mullick, Datta, & Das, 2019) and intra-class (He, Bai, Garcia, & Li, 2008). Intra-class data augmentation methods repeat the original tail class samples without introducing other class knowledge. These methods use samples with low diversity and almost similar contextual information to obtain new sample images (Ahn, Ko, & Yun, 2023). On the contrary,

* Corresponding author at: School of Computer Science, Minnan Normal University, Zhangzhou, Fujian, 363000, China.
E-mail addresses: wenweihecn@163.com (W. He), xujunyan1009@163.com (J. Xu), shijie_jay@126.com (J. Shi), hongzhaocn@163.com (H. Zhao).

https://doi.org/10.1016/j.eswa.2024.123138
Received 7 October 2023; Received in revised form 29 November 2023; Accepted 2 January 2024
Available online 4 January 2024



(a) ECS-based on texture features (b) ECS-based on shape features

Fig. 1. Typical easily confused classes.

inter-class data augmentation methods use the head information to help tail classes generate new samples to expand the diversity of tail samples (Liu, Sun, Han, Dou, & Li, 2020). For example, Park et al. Park, Hong, Heo, Yun, and Choi (2022) conceived a context-rich minority over-sampling strategy.

However, previous data augmentation methods have several limitations. They often overlook the interrelationships between different classes and fail to consider that semantically related classes confuse the model when distinguishing between them. This phenomenon poses a significant challenge for the model in distinguishing easily confused classes. Easily confused samples (ECS) are ubiquitous and one of the significant issues leading to model misclassification (Yue et al., 2019). Fig. 1 illustrates easily confused classes from texture and shape feature perspectives. Fig. 1(a) demonstrates easily confused classes based on texture because the fur texture features of "Cat", "Tiger", and "Leopard" make it challenging for the model to distinguish between them. Fig. 1(b) shows easily confused classes based on shape, where the model often struggles to differentiate between the shape features of "Motor" and "Bicycle". Multi-granularity knowledge considers the interrelationships between classes and provide adequate semantic relationships to help the model identify ECS.

In this paper, we propose a long-tailed data augmentation method, which introduces prior knowledge to select and combine easily confused samples, thereby enhancing classification performance. First, we utilize multi-granularity knowledge to search easily confused classes with semantic relations and create a class relation matrix to record the class relations. We can accurately identify easily confused classes semantically related to tail classes and enhance the efficiency of searching ECS. Second, we utilize a data augmentation strategy on the selected ECS to obtain new samples by pasting the tail class as a foreground clip on the background of the head class sample. We strictly distinguish the boundary line of head and tail classes in new samples and alleviate the scarcity of tail classes by combining tail class samples with abundant head background knowledge. We exploit the fore-background combination method to create more tail class samples to help the model further distinguish tail ECS.

The key contributions of this paper are summarized:

(1) We propose an ECS-based selection and combination method, which generates tail class samples using multi-grained knowledge to search easily confused classes as background.

(2) The presented method can generate several usable tail class samples without more computational cost and readily applies to other long-tailed classification models.

(3) We have demonstrated the proposed ECS-SC effectiveness through experiments and analysis, where ECS-SC outstrips current long-tailed classification methods.

The remainder of this paper is organized as follows. In Section 2, we review the related work. We elaborate on the proposed ECS-SC model in Section 3. Section 4 discusses experimental settings and results analysis. Finally, Section 5 presents the conclusions and future work.

2. Related work

In this section, we offer a concise overview of long-tailed classification methods for re-balancing and data augmentation methods.

2.1. Re-balancing methods

Class imbalance issues often affect the training of classical classifiers, resulting in a bias towards the head class (He & Garcia, 2009). Re-balancing methods have been introduced to address the imbalance phenomenon (Yang, Jiang, Song, & Guo, 2022), including the re-weighting and re-sampling techniques. Re-weighting techniques seek to assign different weights to regulate importance either at the class or instance level. Cui, Jia, Lin, Song, and Belongie (2019) proposed a method to regulate the loss function based on the sample proportion. Similarly, Park et al. (2021) designed a method to calculate the influence weight of similar and heterogeneous samples to adjust the attention of the model. Unlike adjusting sample weights, Cao, Wei, Gaidon, Arechiga, and Ma (2019) utilized the idea of class spacing and adjusted the constraint on the distance between classes to expand the entire class distribution space. Furthermore, Hong et al. (2021) calculated the intra-class and global impact factors to relax the tail boundary. However, re-weighting techniques are often ineffective in large-scale scenarios involving long-tailed distribution, leading to optimization difficulties.

Re-sampling techniques solve the re-weighting optimization difficulties for long-tailed classification, including the under-sampling (Zhao & Zhao, 2024) and over-sampling strategies (Buda, Maki, & Mazurowski, 2018). The under-sampling strategy discards sample data of the majority class to balance long-tailed datasets. For instance, Deng et al. (2016) introduced an under-sampling algorithm that relies on automated clustering to mitigate dataset imbalance. Kang, Chen, and Li (2016) presented a noise filter under-sampling method. Unlike the under-sampling strategy, the over-sampling strategy repeatedly samples more tail samples to achieve a balanced dataset instead of discarding them. Kim, Jeong, and Shin (2020) introduced a technique for generating tail samples, leveraging the ample head class samples to augment the tail class. Moreover, Zhou, Cui, Wei, and Chen (2020) designed the over-sampling method of bilateral-branch sampling further to strengthen the feature representation learning of tail samples. Additionally, Wei, Tao, Xie, Feng, and An (2022) proposed a class-prior method that utilizes open-set noise labels to rebalance the training dataset.

The abovementioned approaches mitigate the long-tailed distribution imbalance in image classification by developing sampling strategies. These re-balancing methods demonstrate promising results, but the long-tailed data still have limitations regarding the tail intra-class diversity and the expansion of its feature space.

2.2. Data augmentation methods

Numerous investigators have explored data augmentation methods to extend the tail intra-class diversity, categorized into intra-class and inter-class data augmentation methods. Intra-class data augmentation methods use simple techniques such as rotation and cropping to generate new samples of their class. He et al. (2008) propounded a simple method to combine samples to generate other samples linearly. Similarly, Yun et al. (2019) linearly superimposed two samples with different probability distributions to expand the data samples. Nonetheless, the convergence speed of the abovementioned methods is significantly slower when adding the same samples. Mullick et al. (2019) introduced generative adversarial minority over-sampling to address the convergence speed issue. However, intra-class data augmentation methods assume long-tailed classes are independent, which results in the model training samples with limited diversity and nearly identical contextual information repeatedly.

The inter-class data augmentation method leverages the head information to assist tail classes in generating new samples, thereby enhancing diversity among tail samples. For instance, Li et al. (2021) aimed to uncover an optimal method for matching covariance matrices to summarize tail samples. Correspondingly, Wang, Huang, Wang, and Leng (2023) designed a deep feature data enhancement method based on guided balancing of semantic relationships. Similarly, Du et al. (2023) introduced a data augmentation method that utilizes cosine similarity to minimize global mixup and local cutmix strategies. However, these methods ignore the deficiency of tail samples and cannot accurately calculate the correlation relationship of tail classes. Due to the abundance of head class samples, recent research introduced the transfer of the head knowledge to the tail class. Liu et al. (2020) presented an approach to aggregate information from each class and expanded the tail class. Additionally, Park et al. (2022) conceived a method of using head class samples to help tail class samples, using the context-rich minority over-sampling. Correspondingly, Ahn et al. (2023) proposed a method to find the data enhancement strength of each class to guide performance improvement.

Existing inter-class data augmentation methods only increase the tail class sample number but ignore the semantic correlation between different classes. This phenomenon causes the model to encounter difficulties distinguishing ECS, thereby reducing the accuracy of long-tail classification. Unlike the approaches discussed earlier, our method uses multi-granularity knowledge to select ECS and then obtains new samples by clipping and combining them. We utilize these obtained samples to train the model and improve its classification performance.

3. The ECS-SC model

In this section, we describe the easily confused sample selection and combination (ECS-SC) based data augmentation method on long-tailed classification.

3.1. Framework overview

We consider the head and tail classes correlation, inspired by the fact that humans in the real-world always associate similar things together (Zhu, Zhu, Wang, Zhang, & Zhao, 2022). We have developed a long-tailed classification method based on data augmentation, which involves selecting and combining easily confused samples (ECS) to expand the tail class diversity. The basic model diagram of ECS-SC is shown in Fig. 2.

(1) **Easily confused samples selection (ECS-S):** The sampling network of ECS-SC consists of uniform and reverse samplers. We explore the inter-class correlations and capture them using a class relationship matrix by leveraging multi-granularity knowledge. Furthermore, we exploit this information to select easily confused tail samples from the bilateral branch samplers (BBS).

(2) **Easily confused samples combination (ECS-C):** We combine ECS utilizing the data augmentation method. We follow the basic procedure of the CutMix method (Yun et al., 2019) by cropping data-sparse tail sample images as the foreground and content-rich head samples as the background. Then, we splice and combine these two parts into a new sample.

(3) **Long-tailed classification:** We introduce the whole process of long-tailed model training, including feature representation learning, loss function, and prediction. Moreover, we provide a detailed explanation of the ECS-SC process and the classification process of the model.

3.2. Easily confused sample selection (ECS-S)

In this section, we process the long-tailed distribution from the sample and class perspectives. From the perspective of the sample, we use BBS, including the uniform and reverse sampler. From the class perspective, we borrow multi-granularity knowledge to build a semantic relation tree to define and select easily confused samples.

From the perspective of sample, our sampler network consists of uniform and reverse samplers. Given a long-tailed distributed dataset $\mathbf{D} = \{(x_i, y_i)\}_{i=1}^{N}$ of C classes, where N denotes the sample number, and C is the class number. In the uniform sampler, samples from each class have the same probability of being selected. The sampling probability p_i is demonstrated as:

$$p_i = \frac{n_i}{\sum_{j=1}^{C} n_j}, \tag{1}$$

where n_i denote the ith class sample number. Let the sample obtained by the uniform sampler be $\tilde{\mathbf{U}} = \{(x_1^u, y_1^u), \ldots, (x_k^u, y_k^u), \ldots, (x_B^u, y_B^u)\}$, where $\tilde{\mathbf{U}}$ is set of the uniform sampler, and B represents the size of the sampling batch.

In the reverse sampler, the sampling probability is proportional to the reciprocal of the sample number of this class to the total sample size. Then, the reverse sampler probability for the ith class is defined by:

$$p_i = \frac{1/n_i^{\delta}}{\sum_{j=1}^{C} 1/n_j^{\delta}}, \tag{2}$$

where the ith class sampling probability is inversely proportional to n_i^{δ}. Meanwhile, the sampling probability is adjusted by setting the value of δ, and as the value of δ increases, the weights of the tail class increase. Similarly, we assume the sample obtained by the reverse sampler as $\tilde{\mathbf{R}} = \{(x_1^r, y_1^r), \ldots, (x_l^r, y_l^r), \ldots, (x_B^r, y_B^r)\}$, where $\tilde{\mathbf{R}}$ is set of the reverse sampler.

The uniform sampler randomly selects samples with the same probability, preserving the long-tailed originally distributed characteristics. The reverse sampler emphasizes the importance of tail classes and corrects the issue of the model being biased towards the head classes. We combine these two samplers and use them as the sampling network of ECS-SC.

From the perspective of class, we first utilize multi-granularity knowledge to search for semantic relationships between classes. The multi-granularity relations are derived from word vector frequencies of WordNet and semantic information of class labels. We take advantage of the distance between word vectors with strong correlation, much smaller than between word vectors with weak or no correlation. We adopt a distance measurement strategy based on word vector correlation to aggregate fine-grained class word vectors with strong correlation into a larger coarse-grained class. During this process, we also utilize the semantic information of class labels to guide and correct these coarse-grained classes. This strategy helps us cluster classes and provides a way to understand the relationships between classes at multiple levels of granularity.

We use trees as data structures to present these relationships and better represent multi-granularity relationships between classes. By

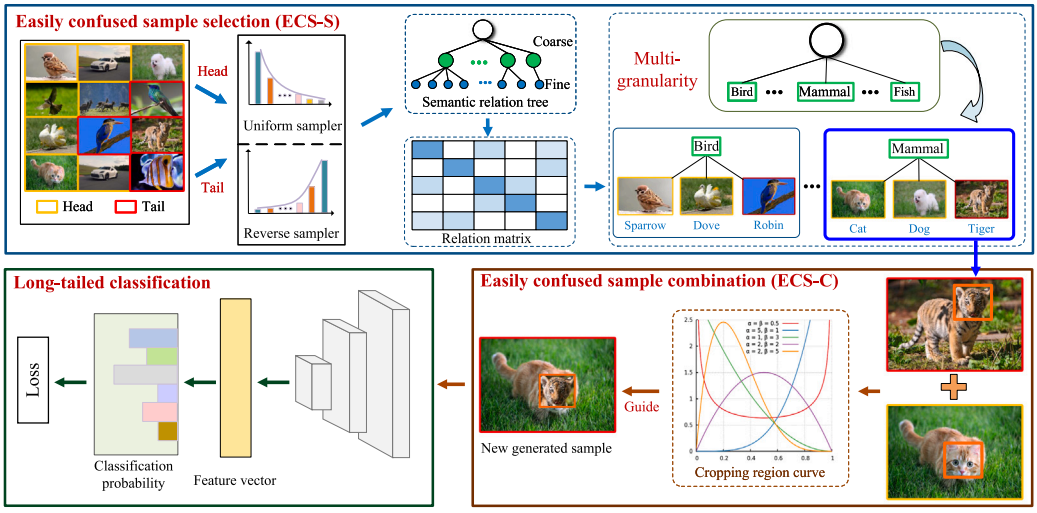


Fig. 2. Framework of the ECS-SC model. The model selects ECS based on the semantic relationship subtree within the sampler network. The relationship matrix records information about the semantic relationships between classes. In ECS-SC, the samples with red and yellow borders represent those sampled from the tail and head classes, respectively. Subsequently, the cropping regions are determined and combined based on the cropping curve to generate additional new samples for model training.

building a semantic relationship tree in WordNet, we search for classes with sibling relationships and define these classes with sibling relationships as easily confused classes. Such a multi-granularity relationship tree improves the understanding of fine-grained classes and provides the model with more in-depth and hierarchical semantic information. We utilize Example 1 to elaborate ECS intuitively.

Example 1. Fig. 3 illustrates our process of searching ECS, which displays a subtree of class semantic relationships, and we select ECS based on class semantic relationships. The coarse-grained class "Felidae" is the parent class of the fine-grained classes "Cat", "Tiger", and "Leopard". They belong to the same sibling set, defined as easily confused classes. Similarly, the coarse-grained class "Vehicle" includes the fine-grained classes "Motor" and "Bicycle", where "Motor" and "Bicycle" are considered to be confusing classes with sibling relationships.

We construct a class relationship matrix **A** to map the relationship between different classes. We flag those classes that are easily confused with other classes in the class relationship matrix, which is expressed as follows:

$$A_{ij} = \begin{cases} 1, & if\ i \in S(j), \\ 0, & otherwise, \end{cases} \tag{3}$$

where $S(j)$ represents the sibling set of the jth class. We use Example 2 to provide a more intuitive explanation of class relationship matrix **A**.

Example 2.

Fig. 4 illustrates the creation process of the class relation matrix. In Fig. 4(a), a semantic relation subtree is generated with multi-granularity relations. Here, r_1 represents the root node of the subtree, and g_1 and g_2 are its two branches. The child node set of g_1 is $\{c_1, c_2, c_3, c_4, c_5\}$, and similarly, the child set of g_2 is $\{c_6, c_7\}$. $\{c_1, c_2, c_3, c_4, c_5\}$ shares the same parent node, indicating a sibling relationship among them. For example, the sibling set of node c_1 is denoted as $S_{c_1} = \{c_2, c_3, c_4, c_5\}$. Similarly, the set $\{c_6, c_7\}$ is also a sibling set, with the sibling set of node c_6 denoted as $S_{c_6} = c_7$. We calculate easily confused classes based on sibling relationships in the tree structure. When there is a sibling relationship between two classes, we assign a value of 1 to the corresponding elements in the class relation matrix. If there is no sibling relationship, the elements are not assigned. Fig. 4(b) shows the class relation matrix constructed based on Fig. 4(a), where different background colors represent different sibling sets.

We obtain the head and tail class samples through the BBS sampler and then randomly select the head class samples that are related to the tail class according to the value of the class relationship matrix **A**. We

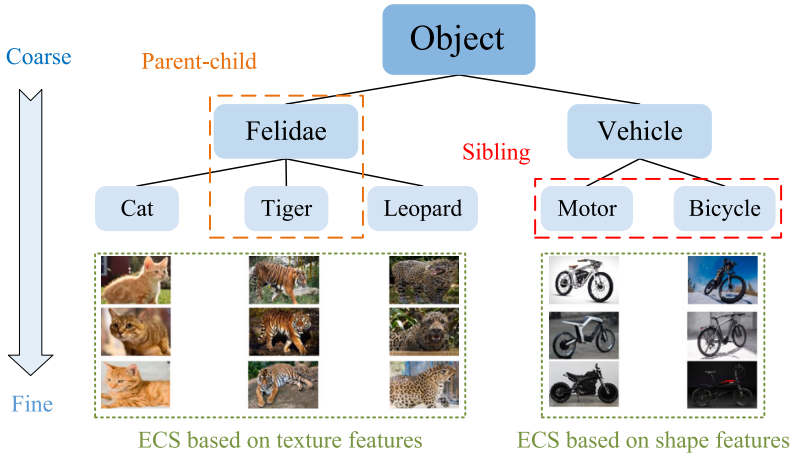
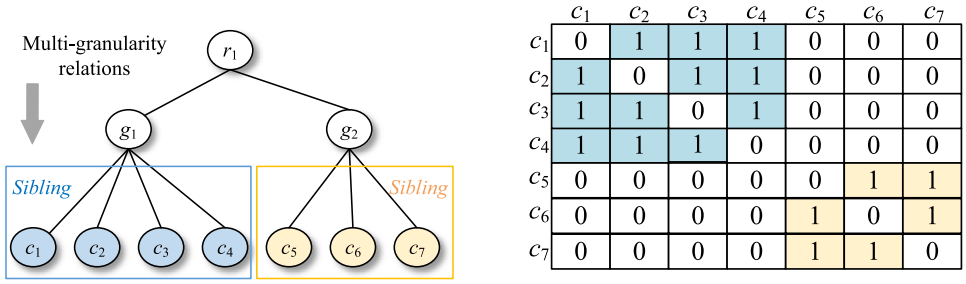


Fig. 3. Semantic relationship subtree from coarse-to-fine.



(a) Semantic relationship subtree

	c_1	c_2	c_3	c_4	c_5	c_6	c_7
c_1	0	1	1	1	0	0	0
c_2	1	0	1	1	0	0	0
c_3	1	1	0	1	0	0	0
c_4	1	1	1	0	0	0	0
c_5	0	0	0	0	0	1	1
c_6	0	0	0	0	1	0	1
c_7	0	0	0	0	1	1	0

(b) Class relationship matrix

Fig. 4. Transformation from semantic relationship tree to class relationship matrix.

achieve a more efficient ECS-S process by introducing the class-relation matrix and BBS. Compared with directly traversing the entire dataset to search ECS, the class relationship matrix can quickly determine which classes have confusing relationships. Meanwhile, the reverse sampler can select samples that are easily confused with tail classes, reducing the complexity and computational overhead of the selection process.

3.3. Easily confused sample combination (ECS-C)

In this section, we introduce the overall process of easily confused sample data augmented combination. We propose ECS-C to obtain an information-rich sample from the background (x_k^u, y_k^u) and foreground sample image (x_l^r, y_l^r), where the image set $\tilde{\mathbf{U}}$ are biased to the head classes, and the image set $\tilde{\mathbf{R}}$ contains more tail class samples.

We first calculate a rectangular marquee $Q_l = (w_e, h_e, w_f, h_f)$ to select images of the foreground tail class. Second, we crop the foreground image x_l^r based on the vertex coordinates provided by Q_l. Finally, the region in the background image x_k^u corresponding to Q_l is discarded and replaced with the image x_l^r. The calculation of the rectangular marquee Q_l is based on:

$$\begin{cases} w_e \sim \text{Unif}\,(0, W), & w_f = W\sqrt{1-\beta}, \\ h_e \sim \text{Unif}\,(0, H), & h_f = H\sqrt{1-\beta}, \end{cases} \tag{4}$$

where W and H represent the image width and height. The rectangular marquee ratio β sampled from beta distribution $\text{Beta}(\alpha, \alpha)$, where we set α to 1 in our experiments. The combination ratio of the foreground and the background image is determined according to the combination coefficient λ. The clipping ratio λ is calculated as follows:

$$\lambda = \frac{(w_f - w_e) \times (h_f - h_e)}{W \times H}, \tag{5}$$

where (w_e, h_e, w_f, h_f) are the vertex coordinates of the clipping region.

ECS-C utilizes head classes with ample intra-class information to complement sample-scarce tail classes. We synthesize the head class samples as the background image and samples from the tail class as the foreground image. The generated training set sample, denoted as $\tilde{x}$, $\tilde{y}$, is utilized to train the long-tailed model. We utilize the combination operation as follows:

$$\begin{cases} \tilde{x_n} = (1 - Q_l) \odot x_k^u + Q_l \odot x_l^r, \\ \tilde{y_n} = (1 - \lambda) y_k^u + \lambda y_l^r, \end{cases} \tag{6}$$

where λ represents the clipping ratio of the foreground tail image, and $\odot$ means element-wise multiplication. ECS-C in training iteration generates a new sample $(\tilde{x_n}, \tilde{y_n})$ with more information-rich by combining the two picked, easily confused head and tail training samples.

3.4. Long-tailed classification

In this section, we introduce an overview of the complete training procedure for the ECS-SC long-tailed classification model.

First, we utilize convolutional neural networks (CNN) and global average pooling (GAP) to learn the sample feature representation during model training. Second, GAP calculates a feature vector for each image. This feature vector is then mapped to the fully connected layer. Finally, the mapped features are input into the softmax layer to predict the class probability distribution. Symbol $f_\varphi(\cdot)$ denotes the feature extraction module, and matrix $\mathbf{W}_{fc}$ is the parameters of fully connected layers. We take the new sample $\tilde{x_n}$ as input to the feature extraction module and then extract sample features $f_\varphi(\tilde{x_n})$. The softmax function maps the sample features to predict the classes and calculates the ith class probability:

$$\hat{p}_i = \frac{e^{\tilde{y}_i}}{\sum_{j=1}^{C} e^{\tilde{y}_j}}, \tag{7}$$

where $\tilde{y}_i$ is the predicted output of the feature vector.

The ECS-SC total loss function $\mathcal{L}$ consists of two parts and is calculated according to Eq. (6). We continuously optimize model parameters by strictly defining the boundaries of ECS, making our proposed model capable of identifying ECS. The weighted cross-entropy classification loss used in ECS-SC is as follows:

$$\mathcal{L} = (1 - \lambda) E(\hat{p}_k^u, y_k^u) + \lambda E(\hat{p}_l^r, y_l^r), \tag{8}$$

where cross-entropy function $E(\hat{p}_k^u, y_k^u)$ represents the head class loss calculation as the background, and $E(\hat{p}_l^r, y_l^r)$ denotes the tail class loss calculation as the foreground. These two loss parts correspond to learning head and tail sample images.

Discussion of the proposed model: Integrating the two components in Sections 3.2 and 3.3, we obtain the complete ECS-SC model for the long-tailed classification. ECS-SC follows the long-tailed data augmentation clipping method of the traditional fore-background combination but differs from other sample selection methods. In particular, ECS-SC utilizes multi-granularity relationships to reflect semantic information between classes. It selects head class samples that are easily confused with tail classes as backgrounds to help tail classes acquire more head class knowledge, thereby expanding the feature diversity of tail classes.

Compared with traditional data augmentation methods, ECS-SC has the following advantages. First, ECS-SC utilizes head-to-tail samples with semantic relationships to replace the tail sample background instead of randomly selecting samples for replacement. Using semantic relationships to replace the background makes the generated samples more consistent with actual real-world scenarios, and the samples obtained by this method can improve the discriminative ability of the model and be more semantically interpretable. Second, ECS-SC does not add more calculations to model training and is a lightweight data augmentation method because it does not increase the number of training samples. Finally, ECS-SC is transferable as it is transplanted to other long-tailed classification methods. Suppose the class information of a dataset is single, and coarse-grained classes cannot be found based on semantic relationships. In that case, ECS-SC is downgraded to the traditional data augmentation long-tailed method. Therefore, ECS-SC can be considered an extension of traditional data augmentation methods.

Algorithm 1 provides the pseudo-code for the model training process. Particularly, the relationship matrix $\mathbf{A}$ is obtained in line 2. Our sampling network is described in lines 5 and 6. Lines 7∼ 9 involve selecting and combining ECS, followed by feature extraction in line 10. Lines 11∼ 13 encompass computing the loss function, and parameter updates are performed through backpropagation in line 14.

Algorithm 1 Long-tailed classification for data augmentation based on easily confused sample inference and combination (ECS-SC)

Input: The training epochs number is E_{max}, and B denotes each batch size. The $f_{\varphi}(\cdot)$ denotes feature extraction, and matrix $\mathbf{W}_{fc}$ is the parameters of fully connected layers. The background image is sourced from the uniform sampler as x_k^u, and the reverse sampler provides the foreground x_l^r, obtaining the new image $\tilde{x_n}$.
Output: The parameter $\mathbf{W}_{fc}$.

1: Initialize parameters $f_{\varphi}(\cdot)$ and $\mathbf{W}_{fc}$;
2: Calculate the class relationship matrix $\mathbf{A}$ according to Eq. (3);
3: **for** $epoch = 1 : E_{max}$ **do**
4: **for** $batch = 1 : B$ **do**
5: Obtain the set $\tilde{\mathbf{U}}$ according to Eq. (1);
6: Obtain the set $\tilde{\mathbf{R}}$ according to Eq. (2);
7: Search easily confused sample from (x_k^u, y_k^u) and (x_l^r, y_l^r);
8: Crop (x_l^r, y_l^r) as foreground according to Eq. (4);
9: Calculate parameter λ according to Eq. (5);
10: Obtain a new sample image $(\tilde{x_n}, \tilde{y_n})$ according to Eq. (6);
11: Extract the new sample image features $f_{\varphi}(\tilde{x})$;
12: Calculate prediction probability $\hat{p}$ according to Eq. (7);
13: Calculate the loss $\mathcal{L}$ according to Eq. (8);
14: Update parameter $\mathbf{W}_{fc}$ by loss $\mathcal{L}$ backpropagation;
15: **end for**
16: **end for**

4. Experimental settings and results analysis

In this section, we detail the experimental setup and analysis of the results of ECS-SC: (1) Experimental datasets and details; (2) Comparison methods; (3) Effectiveness analysis of the BBS strategy; (4) Effectiveness analysis of the ECS-SC strategy; (5) Effectiveness analysis of ECS-SC in local accuracy; (6) ECS-SC comparison with long-tailed methods; and (7) Display of the ECS-SC generated samples.

4.1. Datasets

We employ four datasets in subsequent experiments, consisting of synthetic and real-world long-tailed datasets. These datasets are Cifar-100-LT (He et al., 2016), SUN-LT (Li, Zhao, & Lin, 2022), VOC-LT (Xu, Zhao, & Zhao, 2023), and iNaturalist (Zhao & Zhao, 2024). We utilize the standard version Cifar-100 to construct Cifar-100-LT to ensure the experiment fairness, where the imbalance ratio ρ controls the imbalance degree. The SUN-LT, VOC-LT, and iNaturalist datasets are real-world datasets. Table 1 provides the corresponding information of the datasets.

Table 1
Dataset descriptions.

Dataset	Class	Training set	Imbalance ratio (ρ)
Cifar-100-LT	100	50,000	{10, 50, 100}
VOC-LT	19	2937	49.5
SUN-LT	324	85,147	27.9
iNaturalist	5089	579,184	435.4

All our experiments are performed based on the PyTorch framework. We utilize the ResNet-32 network (He et al., 2016) to train the Cifar-100-LT dataset, and our training strategy for the network for 200 epochs follows (Cao et al., 2019). We train the models on the Ubuntu system with NVIDIA GeForce RTX 2080 Ti and Intel CPU 3.60 GHz. Additionally, we define many-shot classes as those with more than 200 sample images, few-shot classes as those with fewer than 20 sample images, and the remaining classes as medium classes.

4.2. Comparison methods

We compare ECS-SC with several long-tailed methods, and the specific details are as follows:

Comparison with long-tailed methods: (1) CE: Cross-entropy loss (Deboer, Kroese, Mannor, & Rubinstein, 2005); (2) CUDA: Curriculum of data augmentation (Ahn et al., 2023); (3) CB-CE: Class-balanced loss (Cui et al., 2019); (4) BS: Balanced meta-softmax method (Ren, Yu, Ma, Zhao, Yi, et al., 2020); (5) LDAM: Label-distribution-aware margin loss (Cao et al., 2019); (6) OOD: Out-of-distribution method (Wei et al., 2022); (7) M2 m: Major-to-minor method (Kim et al., 2020); (8) Bbn: Bilateral-branch network (Zhou et al., 2020); (9) GHM-CWAP: Intra-and inter method (Zhang, Lin, Zhang, & Xu, 2022); (10) MCKT: Multi-task CNN (Li et al., 2022); (11) CFBSNet: Coarse-to-fine transfer network (Xu et al., 2023); (12) MixSMOTE: Increasing over-sampling diversity method (Xiang, Ding, & Han, 2021). (13) Cutmix: Sample image cropping method (Yun et al., 2019); (14) CMO: Majority class help minority class method (Park et al., 2022); (15) Remix: Re-balanced mixup method (Chou, Chang, Pan, Wei, & Juan, 2020).

4.3. Effectiveness analysis of the BBS strategy

In this section, we compare the accuracy (ACC) of the traditional sampling and BBS strategies. The traditional sampling strategy has the same sampling probability for each class, and the model loss function is calculated based on a single sampler. The BBS strategy consists of uniform and reverse samplers. Fig. 5 shows the sampling results of different samplers. We employ the method of controlling variables to ensure that all other model components remain the same except for the sampler. The results are displayed in Table 2, and the best results are marked in bold.

Table 2
ACC (%) comparison of different sampling strategies on Cifar-100-LT with different ρ.

Imbalance ratio	$\rho = 100$	$\rho = 50$	$\rho = 10$
Single sampler	38.12	42.96	56.45
BBS	**38.61**	**43.30**	**57.11**

From Fig. 5, the following observations are obtained:

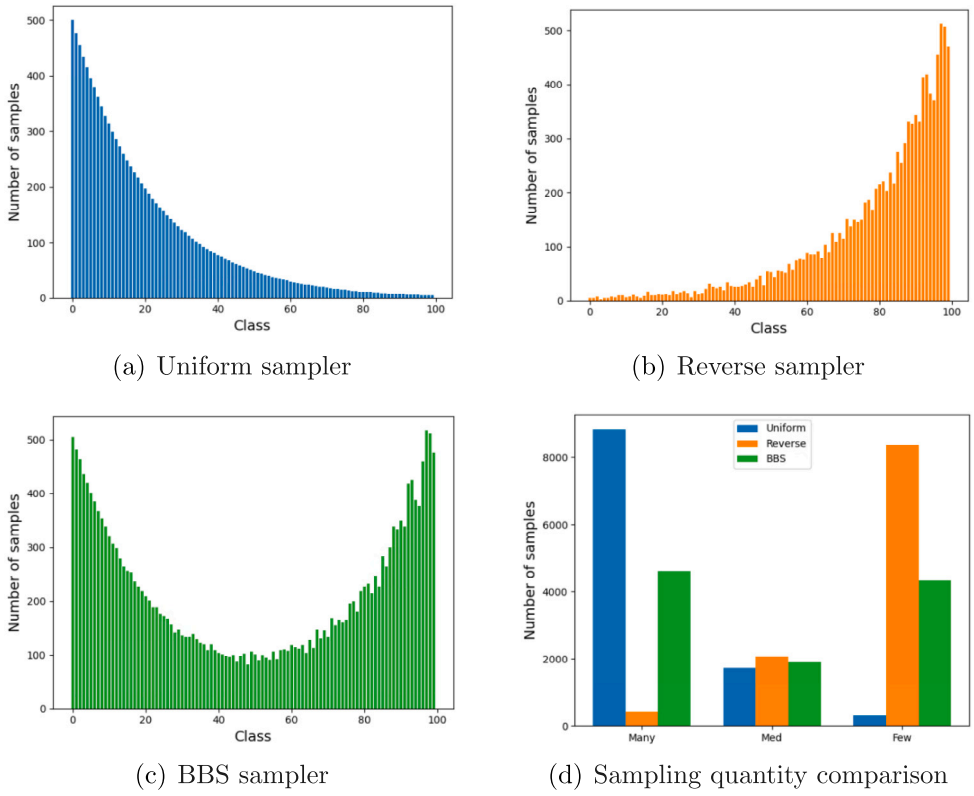
(1) The uniform sampler samples gradually decreased the sampled results from the head to the tail classes according to the original distribution of long-tailed data. The reverse sampler aims to sample more tail class data, and its sampling result is an inverse long-tailed distribution. The uniform sampler results in sparse tail class samples, making the model biased towards the head classes. The reverse sampler increases the number of tail class samples, but it reverses the distribution of long-tailed data and fails to reflect the characteristics of long-tailed data.

(2) The BBS sampler consists of the uniform and reverse samplers and combines both characteristics. From the middle of Fig. 5(c), the first half of the image is in line with the characteristics of long-tailed data distribution, and the second half of the image shows that more tail class samples are sampled, alleviating the bias of model towards the head class. From Fig. 5(d), the number of samples sampled by BBS for many, med, and few classes is between the uniform and reverse samplers, a relatively more balanced state. These show the advantages of BBS sampler. BBS not only retains the original long-tailed data distribution but also helps the model sample more tail class samples, reducing the model excessive tendency towards the head.

The following observations are obtained:

(1) The utilization of BBS yields improved performance on Cifar-100-LT compared with the single sampler strategy. Specifically, BBS achieves a classification accuracy increase of 0.66%, 0.34%, and 0.49% over the CE strategy for imbalance rates of 10, 50, and 100, respectively.

(2) BBS achieves higher accuracy than traditional single sampling at three different imbalance ratios, demonstrating its effectiveness in



(a) Uniform sampler (b) Reverse sampler

(c) BBS sampler (d) Sampling quantity comparison

Fig. 5. Sampling sample details. Fig. 5(a), 5(b), and 5(c) respectively illustrate the sampling results of uniform, reverse, and BBS sampler on the cifar-100 dataset when $\rho = 100$. Fig. 5(d) shows the comparison of the number of samples of the three samplers in many, med, and few classes, respectively.

addressing the long-tailed distribution issue. BBS considers the imbalance issue inherent in long-tailed data and utilizes the reverse sampler to mitigate this problem.

Additionally, the weight of the reverse sampler δ can be adjusted, which helps the model focus on the tail classes. The results of the different BBS weights are presented in Table 3. The best results are marked in bold.

Table 3
ACC (%) comparison of the different reverse sampler weights on Cifar-100-LT ($\rho = 100$).

Weight	Many	Med	Few	All
$\delta = 0.5$	64.37	37.14	8.40	38.28
$\delta = 1$	65.46	**38.57**	**9.66**	**39.58**
$\delta = 2$	65.80	37.11	8.63	38.89
E(δ) (Cui et al., 2019)	**70.40**	38.00	4.72	39.50

We obtain the following observations:

(1) Although raising the value of δ boosts the probability of sampling tail classes, excessively high δ values lead to reduce accuracy. When $\delta = 1$, the ACCs of many, middle, and few classes are improved by 1.09%, 1.43%, and 1.26% compared with $\delta = 0.5$, respectively. Similarly, when $\delta = 1$, the ACCs of middle and few classes are improved by 1.46%, and 1.03% compared with $\delta = 2$, respectively. Therefore, when $\delta = 1$, the model can achieve better performance.

(2) When $\delta = 1$, the reverse sampler outperforms E(δ) with higher classification accuracy on the medium, few, and overall classes, achieving improvements of 0.57%, 4.94%, and 0.08%, respectively. Therefore, after the above analysis, we set the value of δ to 1 in subsequent experiments.

4.4. Effectiveness analysis of the ECS-SC strategy

In this section, we analyze the effectiveness of ECS-SC in enhancing the classification performance of the model. We compare it with the ACC of the CE and CMO methods. Specifically, these methods are divided into forms without any strategies and with sample selection strategies, where CE does not adopt a sample selection strategy. At the same time, CMO and ECS-SC use random and easily confused sample selection strategies, respectively. Table 4 illustrates the experimental results, and the best results are marked in bold.

Table 4
ACC (%) comparison of different sample selection strategies on Cifar-100-LT with different ρ.

Imbalance ratio	$\rho = 100$	$\rho = 50$	$\rho = 10$
CE (Deboer et al., 2005)	37.94	43.13	56.75
CMO (Park et al., 2022)	41.58	47.36	**59.23**
ECS-SC	**43.16**	**47.56**	59.18

We can obtain the following observations:

(1) ECS-S outperforms the traditional CE on Cifar-100-LT with three different ρ. Specifically, the ECS-S performance is improved by 2.43%, 4.43%, and 5.22%, when ρ is 10, 50, and 100, respectively. This indicates that data augmentation methods expand the class feature space, which improves classification accuracy.

(2) ECS-S has better accuracy when $\rho = \{50, 100\}$ compared with CMO. For example, the performance gaps are 0.2% and 1.58%, when $\rho = \{50, 100\}$, respectively. This indicates that the performance gap between CMO and ECS-S becomes more pronounced as ρ increases. Moreover, the easily confused sample selection strategy is effective and can better handle the imbalance issue of long-tailed distribution.

Additionally, we design experiments to prove that the ECS-SC strategy helps distinguish easily confused samples. The coarse-grained class "Human" includes five fine-grained classes on Cifar-100-LT, namely "Baby", "Boy", "Girl", "Men", and "Women". Fig. 6 shows the classification results of the fine-grained class of "Women" in the coarse-grained class of "Human". These five fine-grained classes are easily confused classes, and we represent them with five colors. The left side of Fig. 6 illustrates the prediction results without using ECS-SC in the model, and the right side shows the final prediction results using ECS-SC.

The following observations are obtained:

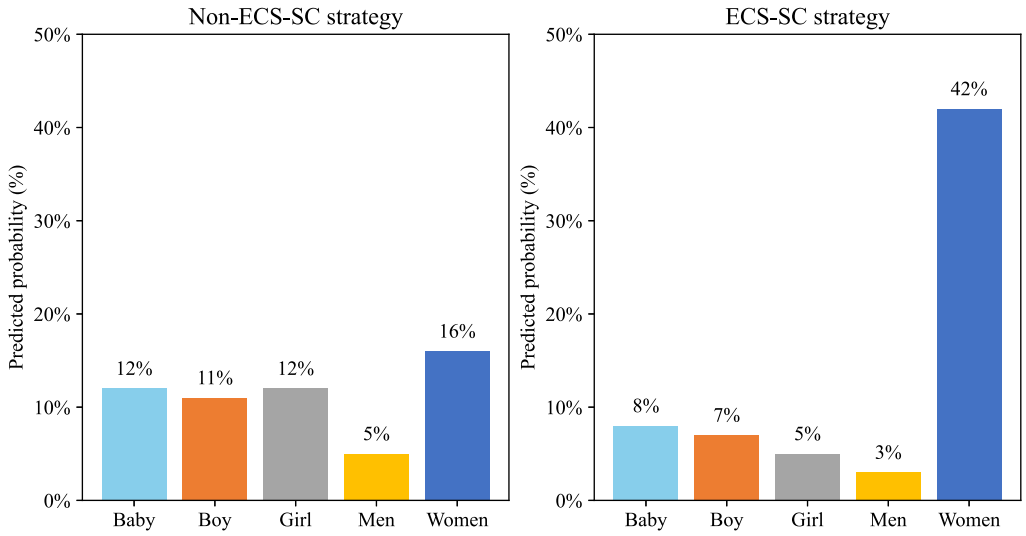


Fig. 6. Woman class specific predicted probabilities.

(1) When the model does not classify "Women" using the ECS-SC strategy, the predicted probability of the "Women" class is lower and is misclassified into its other sibling classes. This means that the predicted probability of the "Women" class is affected by the sibling classes "Baby", "Boy", "Girl", and "Men", making the model difficult to distinguish the "Women" and its easily confused sibling classes.

(2) The prediction probability of the "Women" class is significantly improved after training with the ECS-SC strategy. Compared with the non-ECS-SC prediction of 12%, it has improved to 42%, an increase of 30%. The model reduces the prediction of the "Women" class into its other sibling classes and enhances its classification ability. This result shows the effectiveness of the ECS-SC strategy in handling ECS. Specifically, ECS-SC can reduce the model misclassification predictions of ECS, thereby improving the ability to classify ECS and enhancing the generalization and classification capabilities of the model.

4.5. Effectiveness analysis of ECS-SC in local accuracy

In this section, we analyze the local performance of ECS-SC and specifically demonstrate the classification ACC of the model on each local class. We utilize the Cifar-100-LT dataset and set ρ=100. We take the model as a variable and keep other experimental settings unchanged. CE and LDAM are loss function methods, while CutMix, CMO, CUDA, and ECS-SC are long-tailed data augmentation methods. Symbol * indicates that the data originates from our server, and † indicates that the data comes from the original paper. The experimental results are shown in Table 5, and the best results are marked in bold.

Table 5
ACC (%) comparison of the different sample selection strategies on Cifar-100-LT (ρ = 100).

Method	Many	Med	Few	All
CE* (Deboer et al., 2005)	64.94	36.62	8.50	37.94
CE-DRW* (Cao et al., 2019)	62.49	41.69	15.70	41.25
LDAM* (Cao et al., 2019)	67.40	40.32	7.73	40.03
BS* (Ren et al., 2020)	59.14	41.14	20.93	41.42
CutMix† (Yun et al., 2019)	71.00	37.90	4.90	35.60
CUDA† (Ahn et al., 2023)	70.60	40.34	9.35	41.32
CMO* (Park et al., 2022)	70.14	40.12	10.4	41.58
CE + ECS-SC	69.63	40.37	12.35	43.16
CE-DRW + ECS-SC	61.94	**47.46**	20.89	**45.14**
LDAM + ECS-SC	**72.34**	41.88	9.33	42.45
BS + ECS-SC	51.95	40.65	35.57	43.03
BS-DRW + ECS-SC	52.56	43.23	**36.30**	44.32

We can obtain the following observations:

(1) The accuracies of the many, medium, and few classes in ECS-SC have all improved compared with the traditional CE loss. Specifically, the accuracies of the many, medium, and few classes are 69.63%, 40.37%, and 12.35%, respectively, which are 4.69%, 3.75%, and 3.85% higher than CE, respectively. Similarly, the ECS-SC ACCs of many, medium, and few classes in ECS-SC are improved compared with LDAM loss, which are 2.23%, 0.03% and 4.62% higher than LDAM, respectively. This indicates the effectiveness of ECS-SC, which improves the performance of the many and few classes.

(2) Compared to the CUDA method whose performance is 41.32% on all classes, the ACC of ECS-SC is 43.16%, which is 1.84% better than the CUDA method. ECS-SC is slightly lower than CUDA by 0.97% in many class. In particular, ECS-SC also improves significantly in the few class, exceeding CUDA by 3.00%. This shows that ECS-SC has significant advantages when processing few class samples.

(3) ECS-SC has advantages except for many classes compared with the same type of long-tailed data augmentation methods mentioned above, especially when considering local performance. For example, ECS-SC is slightly lower than CutMix 0.37% and CMO 0.51% in the many classes. Especially, the accuracy of ECS-SC is better than CutMix by 2.47% and CMO by 0.25% in the medium classes, respectively. Similarly, ECS-SC has significantly improved in the few classes, surpassing CutMix by 7.35% and CMO by 1.95%. Moreover, it also outperforms CutMix by 7.56% and CMO by 1.58% in all classes. These results indicate that when compared with the data above augmentation methods of the same type, ECS-SC exhibits more significant performance improvements in the medium and few classes, and selecting easily confused tail samples contributes to improving model performance.

4.6. ECS-SC comparison with long-tailed methods

In this section, we compare ECS-SC with classic and state-of-the-art (SOTA) long-tailed methods. We validate the correctness and feasibility of ECS-SC on several datasets. Symbol * indicates that the data originates from our server, and † indicates that the data comes from the original paper. Typically, a higher imbalance ratio indicates a more challenging long-tailed classification problem. The main classification ACCs of Cifar-100-LT are presented in Table 6.

The following observations are obtained:

(1) Integrating ECS-SC into CE-DRW achieves better results at all imbalance ratios except for ρ = 10, demonstrating its effectiveness and performance. For instance, the accuracy of integrating ECS-SC into CE-DRW is superior to CE-DRW by 2.06% and outperforms Focal by 3.87% when ρ = 10. Its accuracy is superior to CE-DRW by 4.19% and outperforms Focal by 6.54% when ρ = 50. Similarly, the method surpasses CE-DRW by 3.89% and Focal by 6.75% when ρ = 100.

Table 6
ACC (%) comparison of the different ρ on Cifar-100-LT.

Imbalance ratio	$\rho = 100$	$\rho = 50$	$\rho = 10$
CE* (Deboer et al., 2005)	37.94	43.13	56.40
CE-DRW* (Deboer et al., 2005)	41.25	45.60	57.90
Focal† (Lin et al., 2017)	38.39	43.25	56.09
CB-CE† (Cui et al., 2019)	38.60	44.60	57.10
MCKT† (Li et al., 2022)	39.15	44.24	57.15
Mixup† (Wang, Zhao, Wang, Su, & Meng, 2022)	39.54	44.99	58.02
LDAM* (Cao et al., 2019)	40.03	45.01	56.91
OOD† (Wei et al., 2022)	40.26	44.77	58.09
CFBSNet† (Xu et al., 2023)	40.64	46.12	56.87
BS* (Ren et al., 2020)	41.42	46.98	58.39
GHM-CWAP† (Zhang et al., 2022)	41.59	–	57.81
Meta-weight-net† (Shu et al., 2019)	41.61	45.66	58.91
Remix† (Chou et al., 2020)	41.94	-	59.36
LDAM-DRW* (Cao et al., 2019)	42.04	46.20	58.7
M2m† (Kim et al., 2020)	42.16	45.16	57.55
mixSMOTE† (Xiang et al., 2021)	42.34	–	58.48
Bbn* (Zhou et al., 2020)	42.56	47.12	59.16
CMO* (Park et al., 2022)	41.58	47.26	59.38
CE + ECS-SC	43.16	47.32	59.68
CE-DRW + ECS-SC	**45.14**	**49.79**	59.96
LDAM + ECS-SC	42.45	45.76	55.85
BS + ECS-SC	43.03	48.50	59.79
BS-DRW + ECS-SC	44.32	49.50	**60.99**

This means that our method is an effective technique for improving long-tailed classification.

(2) Integrating ECS-SC into CE attains scores of 59.68% and 43.16% on both Cifar-100-LT when $\rho = \{10, 100\}$, which are 0.52% and 0.6% higher, compared with the Bbn model, respectively. It performs better even when the Cifar-100-LT dataset (ρ=100) becomes more imbalanced. This indicates that ECS-SC improvements become more pronounced as the dataset becomes imbalanced. It also further reveals that distinguishing ECS is beneficial for network training. Hence, ECS-SC has high practicability even in more imbalanced datasets.

(3) Compared with the CMO model, integrating ECS-SC into CE achieves higher scores of 0.3% and 1.58% when $\rho = \{10, 100\}$ on Cifar-100-LT, respectively. This indicates that integrating ECS-SC into CE improves classification performance more significantly than the same type of data augmentation methods above, and improvements become more pronounced as the dataset becomes more imbalanced. This illustrates that the ECS-SC data augmentation method enhances the model to distinguish easily confused samples, increase network training, and improve the generalization ability of the model.

(4) Compared with CE, CE-DRW, LDAM, and BS methods, integrating ECS-SC into their models can achieve better improvement. For example, when $\rho = 100$, integrating ECS-SC into the CE, LDAM, and BS methods increases the ACCs by 5.22%, 2.42%, and 1.61% on the CIFAR-100-LT dataset, respectively. Similarly, when $\rho = 10$, integrating ECS-SC into the BS method improves the ACC by 2.60%. This shows that our method has good transferability and effectiveness, can be successfully integrated into existing loss functions and methods, and can achieve better results.

Table 7 illustrates the experimental results on ImageNet-LT datasets. Classification ACC (%) of ResNet-50 with state-of-the-art methods trained for 90 or 100 epochs.

We can obtain the following observations:

(1) Integrating ECS-SC into CE achieves better results on the Image Net-LT dataset, proving its effectiveness and performance. For instance, the accuracy of integrating ECS-SC into CE is superior to CE by 7.33% and outperforms Decouple-cRT by 1.58% on overall accuracy. Similarly, it surpasses CE by 3.32% and Decouple-cRT by 8.55% at many class accuracies. This means that our method is an effective technique for improving long-tailed classification.

(2) The ACCs of integrating ECS-SC into CE many, medium, and few classes in ECS-SC is improved compared with CMO model, which are

Table 7
ACC (%) comparison of the different long-tailed methods comparison on ImageNet-LT.

	Many	Med	Few	All
CE† (Deboer et al., 2005)	64.00	33.80	5.80	41.60
CE-DRW† (Deboer et al., 2005)	61.70	47.30	28.80	51.10
Decouple-cRT† (Kang et al., 2020)	58.77	44.03	26.12	47.35
Decouple-LWS† (Kang et al., 2020)	57.10	45.20	29.33	47.70
Remix† (Chou et al., 2020)	60.40	46.90	30.70	48.60
CMO* (Park et al., 2022)	67.27	40.70	15.51	47.80
CUDA† (Ahn et al., 2023)	67.20	47.00	13.50	47.32
LDAM* (Cao et al., 2019)	68.45	41.76	13.42	48.07
BS* (Ren et al., 2020)	60.96	48.36	32.56	50.90
CE + ECS-SC	67.32	41.61	23.94	48.93
CE-DRW + ECS-SC	62.49	48.81	25.27	50.90
LDAM + ECS-SC	**68.53**	41.87	19.07	49.66
BS + ECS-SC	61.91	**49.56**	**33.39**	**52.10**

0.05%, 0.91% and 8.43% higher than CMO, respectively. Similarly, its accuracy is superior to CMO by 1.13% on overall accuracy. This indicates the effectiveness of integrating ECS-SC into CE, which improves the performance of the many and few classes.

(3) Compared with CE, LDAM, and BS methods, integrating ECS-SC into their models can achieve better improvement. This underscores the transferability and effectiveness of ECS-SC, demonstrating its successful integration into existing loss functions and methods for achieving enhanced results.

Table 8 illustrates the experimental results on other datasets. Symbol * indicates that the data originates from our server, and † indicates that the data comes from the original paper.

Table 8
ACC (%) comparison of different loss functions on the VOC-LT, SUN-LT, and iNaturalist datasets.

Dataset	SUN-LT	VOC-LT	iNaturalist
CE* (Deboer et al., 2005)	36.94	40.38	7.08
Focal* (Lin et al., 2017)	37.33	40.35	7.26
LDAM* (Cao et al., 2019)	32.23	37.92	7.96
LDAM-DRW* (Cao et al., 2019)	34.80	38.09	7.55
MCKT† (Li et al., 2022)	35.49	38.89	7.11
CFBSNet† (Xu et al., 2023)	**39.46**	39.34	7.32
ECS-SC	39.22	**40.75**	**8.03**

The following observations are obtained:

(1) The accuracies of ECS-SC are 2.28% and 6.99% higher than the CE and LDAM loss model on the SUN-LT dataset. However, ECS-SC is lower than the CFBSNet model by 0.24%. The reason is that the dataset tends to be more balanced when ρ is low, making it difficult for ECS-SC to rely on head class information.

(2) The accuracies of ECS-SC are 0.37% and 2.83% higher than the CE and LDAM loss model on the VOC-LT dataset. Additionally, ECS-SC is higher than the CFBSNet model by 1.14%. This indicates that ECS-SC performs better on datasets with larger imbalance ratios.

(3) The imbalance ratios of SUN-LT, VOC-LT, and iNaturalist are approximately 27.9, 49.5, and 435.4, respectively. ECS-SC achieves the top performance on the VOC-LT and iNaturalist datasets, with higher imbalance ratios. This indicates that ECS-SC is more effective when handling datasets characterized by higher levels of class imbalance.

4.7. Display of the ECS-SC generated samples

In this section, we conduct visualization experiments to more intuitively demonstrate that the ECS-SC method generates final samples by augmenting ECS. Our approach is to increase the feature diversity of tail classes with sparse samples. We utilize the adequate background knowledge in the head classes, which is easily confused with the tail classes, to generate more samples, as shown in Fig. 7. For example, due to the sparse number of samples of the tail class “Panthera tigris”, we utilize the background information of the head class, such as “cat” and

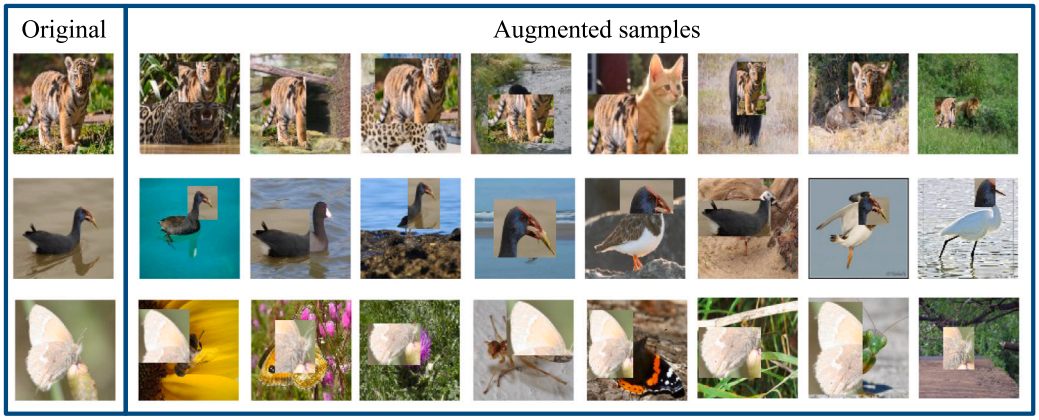


Fig. 7. Display of the ECS-SC generated samples. "Original" represents the original training samples. "Augmented samples" represents visualizations of generated samples via ECS-SC from the corresponding original samples.

"leopard", based on semantic correlation to generate more samples containing the feature information of "Panthera tigris". Similarly, ECS-SC has the same data augmentation strategy for "bird" in the second row and "butterfly" in the third row, which relies on sufficient head class knowledge to expand the diversity of tail class samples. Various head and tail ECS combinations improve the model to learn the generalized and robust representation of the tail classes.

5. Conclusions and future work

In this paper, we propose a long-tailed data augmentation method based on easily confused sample selection and combination (ECS-SC). ECS-SC can enhance the ability of the model to discriminate between easily confused samples in long-tailed classification. The core concept of ECS-SC is rooted in multi-granularity knowledge, which involves selecting easily confused long-tailed data and utilizing abundant head class samples to enrich the diversity of tail class samples. Unlike the traditional long-tailed data augmentation method, ECS-SC presents easily confused sample selection (ECS-S) and easily confused sample combination (ECS-C) strategies. ECS-S utilizes semantic relationships to identify easily confused head and tail class samples in long-tailed data. ECS-C combines these samples foreground–background, boosting the diversity of tail class knowledge by sharing ample head class knowledge. In the experiment, ECS-SC promotes long-tailed classification performance, outperforming some advanced and previous algorithms and data augmentation strategies. Future work will select easily confused samples from more perspectives, not just limited to the semantic relationships of classes.

CRediT authorship contribution statement

Wenwei He: Conceptualization, Methodology, Software, Writing – original draft. **Junyan Xu:** Investigation, Validation, Methodology, Writing. **Jie Shi:** Investigation, Validation, Methodology, Writing. **Hong Zhao:** Conceptualization, Methodology, Validation, Writing – review & editing, Supervision.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 62376114 and the Natural Science Foundation of Fujian Province, China under Grant No. 2021J011003.

References

Ahn, S., Ko, J., & Yun, S. (2023). CUDA: Curriculum of data augmentation for long-tailed recognition. In *International conference on learning representations* (pp. 1–16).

Buda, M., Maki, A., & Mazurowski, M. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, *106*, 249–259.

Cao, K., Wei, C., Gaidon, A., Arechiga, N., & Ma, T. (2019). Learning imbalanced datasets with label-distribution-aware margin loss. In *Advances in neural information processing systems*: *vol. 32*, (pp. 1–12).

Chawla, N., Bowyer, K., Hall, L., & Kegelmeyer, W. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Chen, Z., Fu, L., Yao, J., Guo, W., Plant, C., & Wang, S. (2023). Learnable graph convolutional network and feature fusion for multi-view learning. *Information Fusion*, *95*, 109–119.

Chou, H., Chang, S., Pan, J., Wei, W., & Juan, D. (2020). Remix: Rebalanced mixup. In *European conference on computer vision* (pp. 95–110).

Cui, Y., Jia, M., Lin, T., Song, Y., & Belongie, S. (2019). Class-balanced loss based on effective number of samples. In *IEEE conference on computer vision and pattern recognition* (pp. 9268–9277).

Deboer, P., Kroese, D., Mannor, S., & Rubinstein, R. (2005). A tutorial on the cross-entropy method. *Annals of Operations Research*, *134*, 19–67.

Deng, X., Zhong, W., Ren, J., Zeng, D., & Zhang, H. (2016). An imbalanced data classification method based on automatic clustering under-sampling. In *International performance computing and communications conference* (pp. 1–8).

Du, F., Yang, P., Jia, Q., Nan, F., Chen, X., & Yang, Y. (2023). Global and local mixture consistency cumulative learning for long-tailed visual recognitions. In *IEEE conference on computer vision and pattern recognition* (pp. 15814–15823).

He, H., Bai, Y., Garcia, E., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In *IEEE international joint conference on neural networks* (pp. 1322–1328).

He, H., & Garcia, E. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*, 1263–1284.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *IEEE conference on computer vision and pattern recognition* (pp. 770–778).

Hong, Y., Han, S., Choi, K., Seo, S., Kim, B., & Chang, B. (2021). Disentangling label distribution for long-tailed visual recognition. In *IEEE conference on computer vision and pattern recognition* (pp. 6626–6636).

Kang, Q., Chen, X., & Li, S. (2016). A noise-filtered under-sampling scheme for imbalanced classification. *IEEE Transactions on Cybernetics*, *47*, 4263–4274.

Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J., et al. (2020). Decoupling representation and classifier for long-tailed recognition. In *International conference on learning representations* (pp. 1–16).

Kim, J., Jeong, J., & Shin, J. (2020). M2m: Imbalanced classification via major-to-minor translation. In *IEEE conference on computer vision and pattern recognition* (pp. 13896–13905).

Krizhevsky, A., Sutskever, I., & Hinton, G. (2017). Imagenet classification with deep convolutional neural networks. *Communications of the Association for Computing Machinery*, *60*, 84–90.

Li, S., Gong, K., Liu, C., Wang, Y., Qiao, F., & Cheng, X. (2021). Metasaug: Meta semantic augmentation for long-tailed visual recognition. In *IEEE conference on computer vision and pattern recognition* (pp. 5212–5221).

Li, X., & Shen, Q. (2023). A hybrid framework based on knowledge distillation for explainable disease diagnosis. *Expert Systems with Applications*, Article 121844.

Li, Z., Zhao, H., & Lin, Y. (2022). Multi-task convolutional neural network with coarse-to-fine knowledge transfer for long-tailed classification. *Information Sciences*, *608*, 900–916.

Lin, T., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In *IEEE international conference on computer vision* (pp. 2980–2988).

Liu, J., Sun, Y., Han, C., Dou, Z., & Li, W. (2020). Deep representation learning on long-tailed data: A learnable embedding augmentation perspective. In *IEEE conference on computer vision and pattern recognition* (pp. 2970–2979).

Mullick, S., Datta, S., & Das, S. (2019). Generative adversarial minority oversampling. In *IEEE/CVF international conference on computer vision* (pp. 1695–1704).

Park, S., Hong, Y., Heo, B., Yun, S., & Choi, J. (2022). The majority can help the minority: Context-rich minority oversampling for long-tailed classification. In *IEEE conference on computer vision and pattern recognition* (pp. 6877–6886).

Park, S., Lim, J., Jeon, Y., & Choi, J. (2021). Influence-balanced loss for imbalanced visual classification. In *IEEE/CVF international conference on computer vision* (pp. 735–744).

Patel, A., Jana, S., & Mahanta, J. (2023). Construction of similarity measure for intuitionistic fuzzy sets and its application in face recognition and software quality evaluation. *Expert Systems with Applications*, Article 121491.

Ren, J., Yu, C., Ma, X., Zhao, H., Yi, S., et al. (2020). Balanced meta-softmax for long-tailed visual recognition. In *Advances in neural information processing systems*: *vol. 33*, (pp. 4175–4186).

Shu, J., Xie, Q., Yi, L., Zhao, Q., Zhou, S., Xu, Z., et al. (2019). Meta-weight-net: Learning an explicit mapping for sample weighting. In *Advances in Neural Information Processing Systems*: *vol. 32*, (pp. 1–12).

Varghese, A., Jawahar, M., & Prince, A. A. (2023). Learning species-definite features from digital microscopic leather images. *Expert Systems with Applications*, *224*, Article 119971.

Wang, Y., Huang, E., Wang, R., & Leng, T. (2023). Bsda in visual recognition: Balanced semantic data augmentation for long-tailed data. In *International joint conference on neural networks* (pp. 1–8).

Wang, Y., Liu, R., Lin, D., Chen, D., Li, P., Hu, Q., et al. (2023). Coarse-to-fine: Progressive knowledge transfer-based multitask convolutional neural network for intelligent large-scale fault diagnosis. *IEEE Transactions on Neural Networks and Learning Systems*, *34*, 761–774.

Wang, W., Zhao, Z., Wang, P., Su, F., & Meng, H. (2022). Attentive feature augmentation for long-tailed visual recognition. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*, 5803–5816.

Wang, M., Zhou, L., Li, Q., & an Zhang, A. (2023). Open world long-tailed data classification through active distribution optimization. *Expert Systems with Applications*, *213*, Article 119054.

Wei, H., Tao, L., Xie, R., Feng, L., & An, B. (2022). Open-sampling: Exploring out-of-distribution data for re-balancing long-tailed datasets. In *International conference on machine learning* (pp. 23615–23630).

Xiang, L., Ding, G., & Han, J. (2021). Increasing oversampling diversity for long-tailed visual recognition. In *Artificial Intelligence* (pp. 39–50).

Xu, J., Zhao, W., & Zhao, H. (2023). Coarse-to-fine knowledge transfer based long-tailed classification via bilateral-sampling network. *International Journal of Machine Learning and Cybernetics*, *14*, 3323–3336.

Xue, P., Fei, L., & Ding, W. (2024). A volunteer allocation optimization model in response to major natural disasters based on improved Dempster–Shafer theory. *Expert Systems with Applications*, *236*, Article 121285.

Yang, L., Jiang, H., Song, Q., & Guo, J. (2022). A survey on long-tailed visual recognition. *International Journal of Computer Vision*, *130*, 1837–1872.

Yue, K., Yang, L., Li, R., Hu, W., Zhang, F., & Li, W. (2019). Treeunet: Adaptive tree convolutional neural networks for subdecimeter aerial image segmentation. *International Society for Photogrammetry and Remote Sensing*, *156*, 1–13.

Yun, S., Han, D., Oh, S., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In *IEEE/CVF international conference on computer vision* (pp. 6023–6032).

Zhang, R., Lin, T., Zhang, R., & Xu, Y. (2022). Solving the long-tailed problem via intra-and inter-category balance. In *IEEE international conference on acoustics, speech and signal processing* (pp. 2355–2359).

Zhao, W., & Zhao, H. (2024). Hierarchical long-tailed classification based on multi-granularity knowledge transfer driven by multi-scale feature fusion. *Pattern Recognition*, *145*, Article 109842.

Zhou, B., Cui, Q., Wei, X., & Chen, Z. (2020). Bbn: Bilateral-branch network with cumulative learning for long-tailed visual recognition. In *IEEE conference on computer vision and pattern recognition* (pp. 9719–9728).

Zhu, P., Zhu, Z., Wang, Y., Zhang, J., & Zhao, S. (2022). Multi-granularity episodic contrastive learning for few-shot learning. *Pattern Recognition*, *131*, Article 108820.