Knowledge-Based Systems 311 (2025) 113061

Contents lists available at ScienceDirect

Knowledge-Based Systems

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

DPA-EI: Long-tailed classification by dual progressive augmentation from explicit and implicit perspectives

Yan Zhao, Wenwei He, 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
Semantic learning

ABSTRACT

Deep convolutional neural networks have significantly advanced visual recognition tasks; however, handling long-tailed distribution data remains a substantial challenge for machine learning. Data augmentation is widely recognized as an effective technique for long-tailed image classification. However, most explicit and implicit data augmentation methods fail to balance sample-level generalization with feature-level semantic richness. In this paper, we present a dual progressive augmentation framework designed to mitigate category imbalance in long-tailed classification scenarios. Our framework progressively enhances data diversity and representation by combining explicit and implicit augmentation stages. In the explicit augmentation stage, we employ Mixup training to improve sample-level generalization, thereby increasing the diversity of tail-category data representations. Subsequently, the implicit augmentation stage incorporates category-level covariance learning and meta-learning methodologies, enriching feature-level semantic diversity. Extensive experimental results validate the effectiveness of our method, particularly for tail categories. The proposed framework improves the performance of tail categories by 1%~2% over the second-best method while maintaining the performance of head categories across different long-tailed datasets. Code is available at https://github.com/fhqxa/DPA-EI.

1. Introduction

Deep convolutional neural networks have demonstrated remarkable success in image classification and other computer vision tasks, largely owing to the availability of large-scale datasets such as ImageNet [1] and MS COCO [2]. However, training these models exclusively on long-tailed datasets consistently leads to degraded performance. Specifically, the models exhibit poor generalization in recognizing minority categories and are prone to bias toward majority categories. The long-tailed distribution presents significant challenges, inspiring extensive research across various visual recognition domains, including facial recognition [3], biomedical applications [4], and fault diagnosis [5].

Image classification with long-tailed distributions, as one of the most fundamental tasks in visual recognition, has garnered increasing research attention. Efforts to address this challenge primarily fall into two categories: Re-balancing and data augmentation methods. Re-balancing is a powerful approach to mitigating dataset imbalances, and it can be further divided into two strategies: re-sampling [6] and re-weighting [7,8]. Specifically, re-sampling methods promote training fairness by repeatedly sampling images to balance the dataset. In contrast, re-weighting methods [8] adjust the loss function by assigning higher penalties to tail categories, thereby emphasizing their contribution during training. While re-weighting methods have alleviated data imbalance to some extent, they often lead to overfitting, which is particularly detrimental to tail categories. The core challenge stems from the limited intra-category diversity in tail categories, which restricts their generalization ability on long-tailed datasets.

Data augmentation [9–12] mitigates overfitting by enhancing minority categories, and it can be broadly classified into Explicit Data Augmentation (EDA) and Implicit Data Augmentation (IDA). EDA primarily operates at the pixel level. Specifically, traditional data augmentation methods such as flipping, rotation, translation, and cropping generate new diverse samples for model input. Another approach involves generating new samples by linearly combining existing ones. For instance, Park et al. [13] leveraged the abundant information from majority samples to synthesize new minority samples. Unlike EDA, IDA focuses on the feature level, augmenting data by modifying semantics rather than increasing the quantity of samples. For example, Wang et al. [14] augmented data by searching for meaningful semantic transformations in deep feature space. Similarly, Chen et al. [15] enhanced tail instances by transferring semantic transformation directions from similar categories, thereby increasing intra-category diversity.

* Corresponding author at: School of Computer Science, Minnan Normal University, Zhangzhou, Fujian, 363000, China.
E-mail addresses: g2023062010@stu.mnnu.edu.cn (Y. Zhao), g2022062005@stu.mnnu.edu.cn (W. He), zh1127@mnnu.edu.cn (H. Zhao).

https://doi.org/10.1016/j.knosys.2025.113061
Received 15 November 2024; Received in revised form 17 January 2025; Accepted 19 January 2025
Available online 28 January 2025

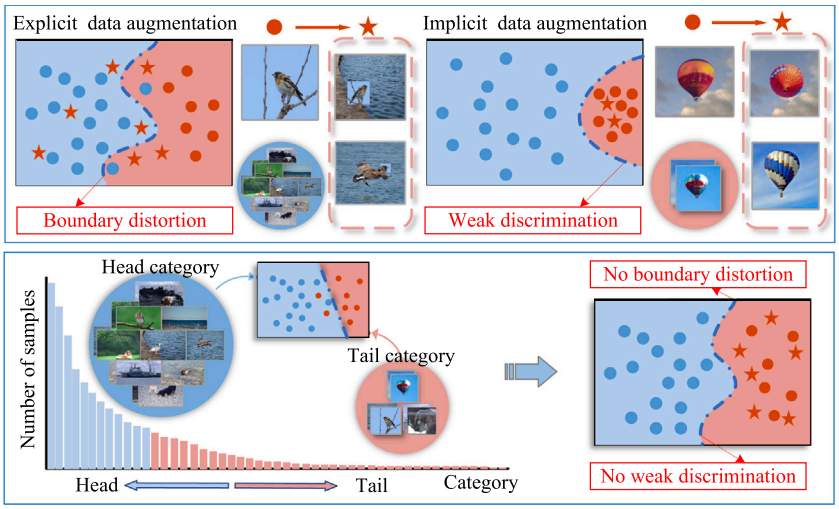


Fig. 1. Illustration of boundary distortion and weak discrimination. In explicit data augmentation, the newly generated images incorporate information from both head and tail categories, which can lead to misclassification-referred to as boundary distortion. In implicit data augmentation, tail samples fail to generate enough new samples, resulting in the persistence of the long-tailed problem-referred to as weak discrimination.

However, using EDA and IDA independently exposes certain limitations. EDA leverages the abundant information from head samples to generate diverse tail samples, which helps improve model generalization. However, the newly generated images often lack interpretability. Additionally, these methods rely heavily on head samples, which can lead to the issue of "boundary distortion." IDA leverages category-specific information to construct an appropriate covariance matrix for meaningful semantic augmentation, enhancing the interpretability of the images. However, it faces the challenge of "weak discrimination" owing to the limited availability of training samples from tail categories. An intuitive explanation of these limitations is provided in Fig. 1. For instance, in EDA, using the tail category "bird" as an example, the method generates additional diverse "bird" samples by utilizing information from head categories. In EDA, the background information from head categories constitutes a significant proportion, often causing the newly generated samples to be misclassified as belonging to the head category. In IDA, using the tail category "balloon" as an example, the limited number of "balloon" samples constrains the semantic directions available, resulting in a relatively small set of features being generated.

In this paper, we propose a data augmentation strategy called Dual Progressive Augmentation from Explicit and Implicit perspectives (DPA-EI) to address category imbalance in long-tailed classification tasks. In the first stage, we utilize the Mixup method [10] to improve the generalization capabilities of the model by mixing samples, thereby enhancing the representation of tail data. In the second stage, we employ a category-level covariance matrix for semantic data augmentation, building on the model trained in the first stage to achieve progressive augmentation. This process involves learning an appropriate category-level covariance from a small, balanced validation set to capture rich semantic information for augmentation. This dual progressive strategy significantly enhances performance, establishing DPA-EI as a prominent and competitive solution for long-tailed classification tasks.

The main contributions of this paper are as follows:

- From a model perspective, we propose a progressive augmentation framework that seamlessly integrates explicit and implicit data augmentation, preserving the representational capacity of the model.
- From a data perspective, we explore semantic information within the data to enable tail categories to acquire more diverse knowledge and enhance their feature diversity.
- Extensive experiments and ablation studies demonstrate the effectiveness of DPA-EI. The method outperforms most existing methods for long-tailed classification.

The remaining sections are organized as follows: Section 2 reviews previous research on long-tailed categorization. Section 3 provides a detailed explanation of the DPA-EI framework. Section 4 presents experimental results and analysis conducted on four benchmarks. Finally, Section 5 concludes the paper and outlines future research directions.

2. Related work

In this section, we provide a concise review of long-tailed classification methods, with a focus on re-balancing and data augmentation strategies.

2.1. Re-balancing methods

Category imbalance [6], particularly in long-tailed datasets, presents a significant challenge in machine learning. To tackle this issue, researchers have developed various re-balancing techniques, including re-sampling and re-weighting.

Re-sampling methods are designed to create a more balanced data distribution through under-sampling and over-sampling techniques. The under-sampling method [16] addresses data imbalance by reducing the number of samples from the majority categories. In contrast, the over-sampling method [17] balances long-tailed datasets by repeatedly sampling additional data from minority categories, without discarding data from the majority categories. For instance, Mahajan et al. [18] rebalanced the training data by calculating a replication factor for each image. Additionally, Kim et al. [19] introduced an optimization technique that transforms prevalent majority categories into corresponding minority categories.

Re-weighting methods [8,20] assign different weights to training samples, which can be applied at the category level or the instance level. The category-level re-weighting method [21] assigns higher weights to samples from tail categories compared with those from head categories, thereby emphasizing the importance of tail categories. In contrast, the instance-level re-weighting method [7] adjusts the weights of individual instances rather than applying global adjustments across entire categories. For example, Lin et al. [22] proposed reducing the weights of easy samples, encouraging the model to focus on harder samples during training. Additionally, recent studies have employed meta-learning techniques to derive weighted loss functions from either a category-level or instance-level perspective. Jamal et al. [23] focused on estimating category weights through meta-learning, enabling a

more balanced weighting across different categories. Conversely, Shu et al. [24] applied meta-learning to dynamically assign instance-wise weights based on gradient direction.

These re-balancing methods demonstrate promising results but struggle to address the lack of diversity in tail categories. Unlike re-balancing methods that primarily aim to balance long-tailed sample distributions, DPA-EI leverages meta-learning to directly enhance feature representation for tail categories, thereby increasing intra-category diversity and expanding the feature space.

2.2. Data augmentation methods

Data augmentation is an effective technique for enhancing intra-category diversity [15], and it can be broadly classified into image-level and feature-level augmentation methods.

Image-level augmentation methods have demonstrated significant promise in visual recognition tasks and have been applied to address imbalanced data problems. For example, in previous studies, CutMix [9] replaced removed regions of an image with patches from another training image, while Mixup [10] linearly interpolated between two images in the training dataset. Follow-up works [25,26] such as Park et al. [13] leveraged CutMix to transfer rich contextual information from majority to minority categories. Zhong et al. [27] proposed a Mixup-based approach to improve representation learning. Similarly, Pan et al. [28] introduced contrastive learning in CutMix and created augmented samples with semantically consistent labels. Additionally, Wang et al. [29] introduced a dynamic optional augmentation strategy, allowing each category to select the most suitable augmentation methods. While image-level augmentation methods have demonstrated satisfactory performance in computer vision tasks, they primarily operate in the raw-pixel image space, which limits their ability to capture deep semantic meaning.

Feature-level augmentation methods [30,31] overcome the limitations of image-level augmentation by leveraging the ability of the model to capture high-level semantic details. These methods modify deep features to create meaningful semantic variations in the image space. For instance, Wang et al. [11] enhanced deep features to guide and balance semantic relationships. Similarly, Liu et al. [32] utilized the angle variance of majority categories to support minority categories, and Chen et al. [26] applied a gaussian distribution model for semantic intra-class transformations. Additionally, recent studies have utilized covariance matrices to identify semantic relationships between categories. For instance, Wang et al. [14] improved deep feature augmentation by incorporating a category-wise conditional covariance matrix. However, this approach was ineffective in estimating a diverse covariance matrix owing to the scarcity of data in minority categories, negatively impacting the accuracy for tail categories. To address this issue, Li et al. [33] utilized meta-learning to derive an optimal category-wise covariance matrix, resulting in more effective augmentation outcomes.

Existing image-level augmentation methods aim to increase the number of samples for tail categories to mitigate overfitting, while feature-level augmentation methods focus on identifying semantic relationships between categories, resulting in notable performance improvements. DPA-EI combines image-level sample augmentation and feature-level semantic augmentation with meta-learning to capture the feature-level covariance matrix. This dual approach allows DPA-EI to operate in both a sample-rich and semantic space, overcoming the limitations of individual image-level and feature-level augmentation methods and offering a complementary solution for data augmentation.

3. The proposed DPA-EI method

This section provides a detailed explanation of the proposed DPA-EI framework for long-tailed categorization.

3.1. Framework overview

Imbalanced training datasets often lead to poor performance on tail categories. To address this issue, we designed a two-part strategy for progressive data augmentation. The basic model structure of DPA-EI is illustrated in Fig. 2.

(1) **Explicit Data Augmentation at the Sample-level (EDAS):** We adapt the standard Mixup method to generate a new training dataset by linearly interpolating between different samples. This approach enhances the representational learning ability of the model during the first stage.

(2) **Implicit Data Augmentation at the Semantic-level (IDAS):** We employ meta-learning to derive appropriate covariance matrices for augmenting minority categories. These covariance matrices encapsulate the semantic transformations of feature directions for each category. As a result, an unlimited number of instances can be generated by applying diverse semantic transformation directions to an instance based on the learned distribution.

(3) **Long-tailed image classification via dual progressive augmentation:** We provide a detailed explanation of the application of explicit and implicit progressive data augmentation strategies, as well as the modifications to the total loss function.

3.2. Explicit data augmentation at the sample-level (EDAS)

In this section, we provide an overview of the EDAS process. Consider the training dataset set $D = \{x_j, y_j\}_{j=1}^{N_u}$ with N_u training samples, where x_j represents the j-th training sample and y_j represents its corresponding category label. The classifier is denoted as $f(\cdot)$, parameterized by the parameters $\boldsymbol{\Theta}$.

The Mixup method generates a new sample $(\tilde{x}, \tilde{y})$ linearly by randomly combining two training samples (x_j, y_j) and (x_l, y_l). The image and label pairs are augmented as:

$$\begin{cases} \tilde{x} = \lambda x_j + (1-\lambda)x_l, \\ \tilde{y} = \lambda y_j + (1-\lambda)y_l, \end{cases} \tag{1}$$

where λ is the mixing coefficient, controlling the extent of image mixing, expressed as:

$$\lambda \sim Beta(\alpha, \alpha), \tag{2}$$

where $Beta(\cdot,\cdot)$ represents beta distribution, and α denotes a hyperparameter derived from it. For EDAS, we define the loss function:

$$\mathcal{L}_{de} = \mathcal{L}_{ce}(f(\tilde{x};\boldsymbol{\Theta}), \tilde{y}), \tag{3}$$

where $\mathcal{L}_{ce}(\cdot,\cdot)$ represents the cross-entropy loss [34], and $f(\tilde{x};\boldsymbol{\Theta})$ denotes the predicted output for the mixed input samples. Fig. 3 illustrates the EDAS process using a straightforward example. In this approach, original and auxiliary samples are combined to generate mixed samples. For example, we use “dog” and “Saxony bird-of-paradise” as representatives of the head and tail categories, respectively. To visually validate this process, we use Grad-CAM [35] to highlight the regions of interest. After the Mixup method is applied, the region of interest in the samples significantly expands. Specifically, the model emphasizes distinctive features, such as the “dog nose” in dog samples and the “head feather” in Saxony bird-of-paradise samples.

3.3. Implicit data augmentation on semantic-level (IDAS)

In this section, we present the IDAS from two perspectives: Implicit semantic data augmentation and meta-learning objective.

Implicit semantic data augmentation. IDA [14] directly augments training data within the feature space by projecting features onto the input space. Furthermore, IDA addresses two key challenges: Determining meaningful semantic transformation directions and avoiding reliance on explicit sampling to obtain these directions.

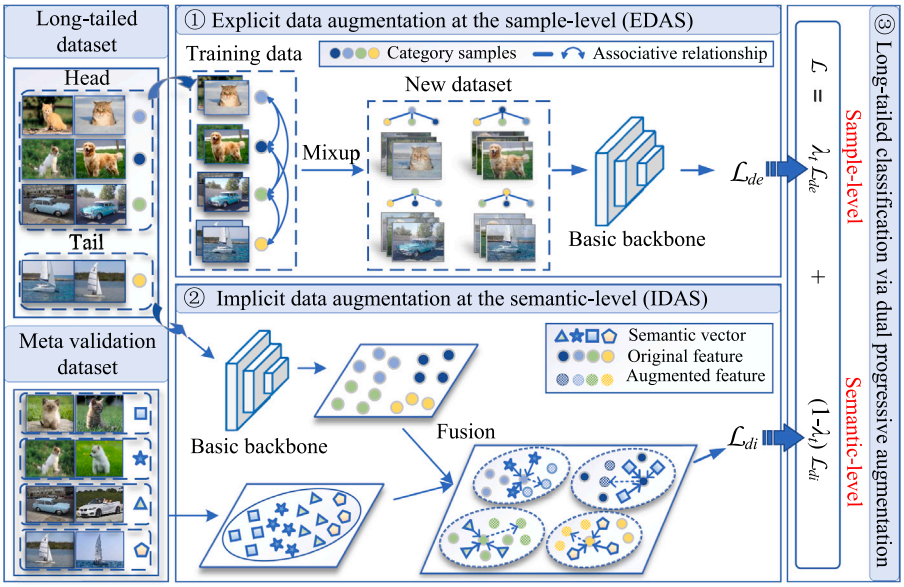


Fig. 2. Framework of the DPA-EI model. The process begins with Mixup training, followed by the next stage of model training to achieve progressive improvements. During the training process, λ_t controls the transition between stages, where $\lambda_t = 1$ corresponds to EDAS, and $\lambda_t = 0$ corresponds to IDAS.

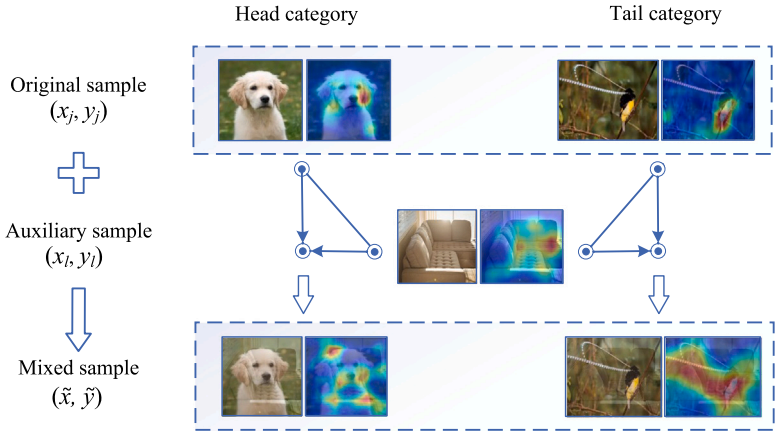


Fig. 3. Illustration of the EDAS process incorporating Grad-CAM.

IDA automatically obtains the semantic transformation direction by sampling random vectors from covariance matrices $\mathbf{\Sigma} = [\mathbf{\Sigma}_1, \ldots, \mathbf{\Sigma}_i, \ldots, \mathbf{\Sigma}_C]$ with C categories and multivariate normal distribution $\mathcal{N}(0, \lambda_s \mathbf{\Sigma}_i)$ with zero mean. Let the feature vector $\mathbf{a} = [\mathbf{a}_1, \ldots, \mathbf{a}_i, \ldots, \mathbf{a}_C]$, where $\mathbf{a}_i$ denotes the feature set of the ith category. The covariance matrices $\mathbf{\Sigma}$ are obtained by summarizing from all mini-batches. Subsequently, the enhanced feature set $\tilde{\mathbf{a}}_i$ are derived by sampling $\mathbf{a}_i$ in the random directions of $\mathcal{N}(0, \lambda_s \mathbf{\Sigma}_i)$. The variable $\tilde{\mathbf{a}}_i$ is expressed as:

$$\tilde{\mathbf{a}}_i \sim \mathcal{N}(\mathbf{a}_i, \lambda_s \mathbf{\Sigma}_i), \tag{4}$$

where λ_s is a positive coefficient that determines the strength of augmentation. Afterward, exploring all possible meaningful directions requires extensive sampling. A simple method involves explicitly sampling each $\mathbf{a}_i$ for M times, thereby creating an enhanced feature set $[\mathbf{a}_i{}^1, \ldots, \mathbf{a}_i{}^k, \ldots, \mathbf{a}_i{}^M]_{i=1}^C$ with a size of $C \times M$. The loss function is represented as:

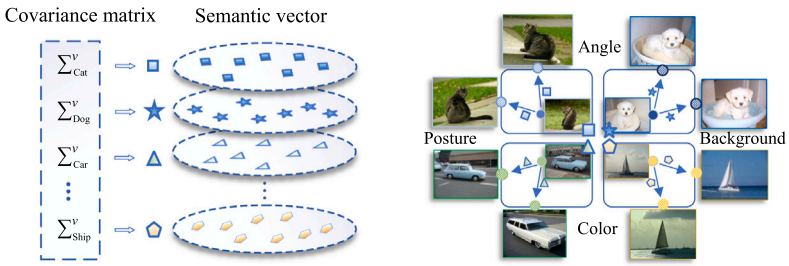
$$\mathcal{L}_{IDA}(\mathbf{W}, \mathbf{b}, \mathbf{\Theta}) = \frac{1}{C} \sum_{i=1}^{C} \frac{1}{M} \sum_{k=1}^{M} -\log \left(\frac{e^{\mathbf{w}_i^T \mathbf{a}_i^k + b_i}}{\sum_{p=1}^{C} e^{\mathbf{w}_p^T \mathbf{a}_i^k + b_p}} \right), \tag{5}$$

where $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_i, \ldots, \mathbf{w}_C]^\top \in \mathbb{R}^C$ and $\mathbf{b} = [b_1, \ldots, b_i, \ldots, b_C]^\top \in \mathbb{R}^C$ represent the weight matrices and biases, respectively. IDA is unable to obtain the exact value of $\mathcal{L}_{IDA}$ when M is large. However, IDA discovers an easily computable upper bound, denoted as $\mathcal{L}_\infty$. Specifically, we replace the computationally infeasible $\mathcal{L}_{IDA}$ with the computable upper bound:

$$\begin{aligned} \mathcal{L}_\infty(\mathbf{W}, \mathbf{b}, \mathbf{\Theta} | \mathbf{\Sigma}) &\geq \underset{M \to \infty}{\mathcal{L}_{IDA}}(\mathbf{W}, \mathbf{b}, \mathbf{\Theta}) \\ &= \frac{1}{C} \sum_{i=1}^{C} -\log \left(\frac{e^{\mathbf{w}_i^T \mathbf{a}_i + b_i}}{\sum_{p=1}^{C} e^{\mathbf{w}_p^T \mathbf{a}_i + b_p + \frac{\lambda_s}{2} (\mathbf{w}_p^T - \mathbf{w}_i^T) \mathbf{\Sigma}_i (\mathbf{w}_p - \mathbf{w}_i)}} \right). \end{aligned} \tag{6}$$

By optimizing this upper bound $\mathcal{L}_\infty$, IDA demonstrates the equivalent semantic augmentation process without explicitly sampling augmented features. In this process, IDA needs to estimate category-wise covariance matrixes $\mathbf{\Sigma}$ for each category. However, it is challenging to sample rich semantic directions for covariance matrices in the tail categories. IDAS integrates a meta-learning component to tackle this issue.

Meta-learning objective. We construct covariance matrices for each category to ensure meaningful semantic transformation directions. Our approach focuses on learning an appropriate category-wise covariance matrix to enhance the performance of tail categories during the augmentation process. The core idea is to leverage meta-learning to guide the construction of the covariance matrix, ensuring that the loss



(a) Visualization of obtaining semantic vector.

(b) Visualization of semantic augmentation.

Fig. 4. The process of category-specific semantic augmentation.

on the validation set is minimized. Specifically, we obtain the optimal category covariance matrix by minimizing the following validation losses on a small validation dataset set $D^v = \{x_j^v, y_j^v\}_{j=1}^{N_u^v}$, where x_j^v represents the j-th validation sample and y_j^v represents its corresponding category label. The category-wise covariance matrix $\mathbf{\Sigma}^v$ is expressed as:

$$\mathbf{\Sigma}^v = \arg\min_{\mathbf{\Sigma}} \sum_{j=1}^{N_u^v} \mathcal{L}_{\text{ce}}(f(x_j^v; \mathbf{\Theta}^*(\mathbf{\Sigma})), y_j^v), \tag{7}$$

where N_u^v represents the total number of validation samples, and $\mathbf{\Theta}^*(\mathbf{\Sigma})$ is the optimal parameter set that depends on the original covariance matrices $\mathbf{\Sigma}$, which is defined as:

$$\mathbf{\Theta}^*(\mathbf{\Sigma}) = \arg\min_{\mathbf{\Theta}} \sum_{j=1}^{N_u} \mathcal{L}_{\infty}(f(x_j; \mathbf{\Theta}), y_j; \mathbf{\Sigma}). \tag{8}$$

We incorporate the new covariance matrices $\mathbf{\Sigma}^v$ to update the upper bound $\mathcal{L}_\infty$. Consequently, the revised $\mathcal{L}_\infty$ is expressed as:

$$\mathcal{L}_{\infty}(\mathbf{W}, \mathbf{b}, \mathbf{\Theta} | \mathbf{\Sigma}^v) = \frac{1}{C} \sum_{i=1}^{C} -\log\left(\frac{e^{\mathbf{w}_i^T \mathbf{a}_i + b_i}}{\sum_{p=1}^{C} e^{\mathbf{w}_p^T \mathbf{a}_i + b_p + \frac{\lambda_s}{2}(\mathbf{w}_p^T - \mathbf{w}_i^T)\mathbf{\Sigma}_i^v(\mathbf{w}_p - \mathbf{w}_i)}} \right). \tag{9}$$

We employ Example 1 to illustrate the IDAS process of semantic data augmentation.

Example 1. In Fig. 4(a), we construct a covariance matrix for each category. These optimized covariance matrices are derived from the validation set, ensuring meaningful semantic transformations. By sampling from these matrices, we generate distinct semantic vectors that capture feature variations within categories. In Fig. 4(b), we use these category-specific semantic vectors to augment the original images of *Cat*, *Dog*, *Car*, and *Ship* and modify attributes such as *Posture*, *Angle*, *Color*, and *Background*. Implicit augmentation introduces subtle but effective semantic adjustments, making the augmented image closer to the original image while enhancing variability.

We train the classifier $f(\cdot)$ by utilizing $\mathcal{L}_\infty$ in Eq. (9), simultaneously completing the process of semantic augmentation. This formula mainly increases the majority of categories instead of the minority categories when directly applying Eq. (9). Therefore, we introduce a re-weighting strategy by setting different weights for different categories to solve this problem. The loss function $\mathcal{L}_{di}$ for IDAS, incorporating the re-weighting strategy, is computed as follows:

$$\mathcal{L}_{di} = \varepsilon_{y_j} \mathcal{L}_{\infty}(f(x_j; \mathbf{\Theta}), y_j; \mathbf{\Sigma}^v), \tag{10}$$

where ε_{y_j} represents the category conditional weight [21], is defined as:

$$\varepsilon_{y_j} \approx \frac{1-\beta}{1-\beta^{n_{y_j}}}, \tag{11}$$

where $\beta = \frac{N_u - 1}{N_u}$ acts as a hyperparameter, and n_{y_j} represents the sample count in the y_j-th category. IDAS places greater emphasis on tail categories, completing the semantic augmentation process for long-tailed datasets.

3.4. Long-tailed image classification via dual progressive augmentation

In this section, we outline the overall process of DPA-EI. First, the classifier is trained using EDAS to enhance its generalization performance. Subsequently, IDAS is employed for semantic augmentation training. DPA-EI has two components, forming the total loss function:

$$\mathcal{L} = \lambda_t \mathcal{L}_{de} + (1 - \lambda_t)\mathcal{L}_{di}, \tag{12}$$

where λ_t regulates the transition from the first stage to the second stage, and its expression is:

$$\lambda_t = \begin{cases} 1 & 0 < t \le T_1, \\ 0 & T_1 < t \le T, \end{cases} \tag{13}$$

where T_1 corresponds to the number of epochs allocated for the first stage, and T corresponds to the total number of epochs allocated for the training process.

Algorithm 1 presents the pseudo-code for the DPA-EI training process. The steps for EDAS are detailed in lines 4∼9: The Mixup coefficient is generated in line 6, mixed samples are created in line 7, and the loss function is calculated in line 8. The steps for IDAS are outlined in lines 12∼18. Line 16 estimates the set of covariance matrices using the validation set.

Algorithm 1 Long-tailed classification by Dual Progressive Augmentation from Explicit and Implicit perspectives (DPA-EI)

Input: Training dataset D, validation dataset D^v, and ending steps T_1, T.
Output: Learned classifier parameters $\mathbf{\Theta}$.

1: Set iteration number $t = 0$;
2: Initialize classifier parameters $\mathbf{\Theta}$;
3: **for** $t \le T_1$ **do**
4: Sample a batch $B = \{(x_j, y_j)\}_{j=1}^{|B|}$ from D;
5: Sample a batch $B^p = \{(x_l, y_l)\}_{l=1}^{|B^p|}$ from D;
6: Obtain Mixup coefficient λ according to Eq. (2);
7: Generate a mixed sample image $(\tilde{x}, \tilde{y})$ according to Eq. (1);
8: Calculate the loss $\mathcal{L}_{de}$ according to Eq. (3);
9: Update model parameters: $\mathbf{\Theta} \leftarrow \mathbf{\Theta} - \alpha \nabla_{\mathbf{\Theta}} \mathcal{L}_{de}$;
10: **end for**
11: **for** $T_1 < t \le T$ **do**
12: Sample a batch $B = \{(x_j, y_j)\}_{j=1}^{|B|}$ from D;
13: Sample a batch $B^v = \{(x_j^v, y_j^v)\}_{j=1}^{|B^v|}$ from D^v;
14: Estimate the covariance matrices $\mathbf{\Sigma} = [\Sigma_1, \cdots, \Sigma_{y_j}, \cdots, \Sigma_C]$;
15: Calculate loss: $\mathcal{L}_{B^v} = \frac{1}{|B^v|} \sum_{j=1}^{|B^v|} \mathcal{L}_{ce}(f(x_j^v; \mathbf{\Theta}(\mathbf{\Sigma})), y_j^v)$;
16: Update covariance matrix $\mathbf{\Sigma}^v$ according to Eq. (7);
17: Calculate the loss $\mathcal{L}_{di}$ with updated $\mathbf{\Sigma}^v$ according to Eq. (10);
18: Update model parameters: $\mathbf{\Theta} \leftarrow \mathbf{\Theta} - \alpha \nabla_{\mathbf{\Theta}} \mathcal{L}_{di}$;
19: **end for**

Table 1
Statistical information of the long-tailed dataset.

Dataset	Category	Training set	Imbalance ratio (ρ)
CIFAR-LT-10	10	50,000	{10, 20, 50, 100, 200}
CIFAR-LT-100	100	50,000	{10, 20, 50, 100, 200}
ImageNet 2012	1000	1,281,167	1280/5
iNaturalist 2018	8142	435,713	1000/2

4. Experimental results and analysis

In this section, we present the experimental setup and analysis of the DPA-EI framework as follows: (1) Experimental configuration; (2) Comparison with existing methodologies; (3) Analysis of local accuracy performance, with focus on many-shot, medium-shot, and few-shot categories; (4) Comparative analysis of DPA-EI with related long-tailed methods; (5) Evaluation of the effectiveness of the DPA-EI strategy; (6) Ablation studies; and (7) Visualization of results.

4.1. Experimental setup

Long-tailed datasets. We assess the performance of DPA-EI on the following long-tailed datasets: CIFAR-LT-10, CIFAR-LT-100 [36], ImageNet 2012 [1], and iNaturalist 2018 [37]. Table 1 presents the relevant details for the datasets.

• **Long-tailed CIFAR.** The CIFAR is categorized into 10 categories for CIFAR-10 and 100 categories for CIFAR-100. We utilize the imbalance ratio ρ to control the imbalance degree. We obtain training sets with different imbalance degrees by varying $\rho \in \{10, 20, 50, 100, 200\}$. We randomly select ten samples from each category to form a validation set D^v.

• **Long-tailed ImageNet.** The ImageNet-LT consists of 1000 categories sourced from ILSVRC2012, comprising 1,281,167 training images and 50,000 validation images. We select ten images per category to construct D^v.

• **Long-tailed iNaturalist.** The iNaturalist dataset is a vast collection of real-world images characterized by an extremely imbalanced category distribution. Therefore, we choose two images per category to create D^v of iNaturalist 2018.

General implementation. The experiments are conducted using the PyTorch framework on an Ubuntu system equipped with an NVIDIA GeForce RTX 2080 Ti GPU and an Intel CPU operating at 3.60 GHz. For comparative analysis, results from original papers conducted under the same settings are directly referenced. Additionally, the dataset is divided into three categories-many-shot, medium-shot, and few-shot-for local accuracy analysis. Specifically, many-shot categories consist of more than 200 samples, few-shot categories contain fewer than 20 samples, and the remaining categories fall into the medium-shot category.

4.2. Comparison methods

To comprehensively evaluate the performance of DPA-EI against existing methods, eight baseline methods from typical categories are introduced, as summarized in Table 2. We compare DPA-EI with the following methods:

• **Cross-entropy method.** Cross-entropy (CE) is a fundamental loss function in the experiment [34].

• **Category-level and instance-level re-weighting methods.** These methods assign weights to training samples, including (1) CB-CE: Category-balanced loss [21]; (2) Focal: Focal loss [22]; and (3) LDAM: Label-distribution-aware margin loss [38].

• **Meta-learning methods.** We compare with methods that utilize meta-learning techniques, including (1) Meta-class-weight: Rethinking category-balanced methods from a domain adaptation perspective [23] and (2) Meta-weight net: Altering sample weighting by explicit mapping [24].

Table 2
Comparison methods.

Category	Model
Cross-entropy training	CE
Category-level re-weighting	CB-CE; Focal
Instance-level re-weighting	LDAM
Meta-learning	Meta-weight net; Meta-class-weight
Image-level data augmentation	CMO; DODA; PLA
Feature-level data augmentation	MetaSAug; HCKC
Two-stage training	BBN; DR-DPM

Table 3
Test accuracy (%) on CIFAR-LT-100 ($\rho = 100$). The method represented by "†" is taken from the results reported in [41]. The highest results are marked in bold.

Method	Many	Med	Few	All
CE [34]	64.94	36.62	8.50	37.94
CE-DRW† [38]	62.49	41.69	15.70	41.25
LDAM† [38]	67.40	40.32	7.73	40.03
CutMix† [9]	71.00	37.90	4.90	35.60
MetaSAug with CE [33]	68.41	47.95	18.68	44.71
CMO† [13]	70.14	40.12	10.40	41.58
CUDA† [42]	70.60	40.34	9.35	41.32
DODA with CE* [29]	**74.80**	43.80	10.00	44.50
ECS-SC with CE† [41]	69.63	40.37	12.35	43.16
DPA-EI (Ours)	67.03	**50.03**	**22.24**	**46.82**

• **Image-level and feature-level data augmentation methods.** We compare with existing data augmentation methods, including (1) HCKC: Hierarchical Convolutional Neural Network with Knowledge Complementation [12]; (2) CMO: Majority category help minority category method [13]; (3) MetaSAug: Meta semantic augmentation [33]; (4) DODA: Dynamic Optional Data Augmentation [29]; and (5) PLA: Pixel-level augmentation [26].

• **Two-stage methods.** These methods adopt two-stage learning, including (1) BBN: Bilateral-branch network [39] and (2) DR-DPM: A dual-phase model to disentangle the overall gradient [40].

4.3. Effectiveness analysis of DPA-EI in local accuracy

We provide classification accuracy for each category-many-shot, medium-shot, few-shot, and all-shot-and compare DPA-EI with basic loss functions and various data augmentation methods. Table 3 presents the experimental results on CIFAR-LT-100 ($\rho = 100$). The key findings are as follows:

(1) DPA-EI shows improved accuracy across all categories compared with CE loss. Specifically, the accuracies are 67.03% for many-shot, 50.03% for medium-shot, and 22.24% for few-shot, reflecting increases of 2.09%, 13.41%, and 13.74%, respectively. Compared with other basic loss function methods, DPA-EI achieves an overall accuracy of 46.82%, with particularly notable improvements in the medium-shot and few-shot categories.

(2) Compared with long-tailed data augmentation methods, DPA-EI demonstrates superior local performance across various categories, except for the many-shot category. For example, DPA-EI is lower than DODA with CE [29] 4.77% in the many-shot, but the accuracy rate is substantially improved by 6.23%, 12.24%, and 2.32% in the medium-shot, few-shot, and all-shot, respectively. These results indicate that DPA-EI is particularly well-suited for handling extremely imbalanced data. Its effectiveness stems from its enhanced ability to improve feature representation and classification performance in less representative categories.

Table 4
Test accuracy (%) of ResNet-32 on CIFAR-LT-10 under different imbalance settings. "*" indicates results reported in the original paper.

Imbalance ratio (ρ)	200	100	50	20	10
CE [34]	67.65	70.35	74.68	81.77	86.12
CB-CE* [21]	68.89	74.57	79.27	84.36	87.49
Focal* [22]	65.29	70.38	76.71	82.76	86.66
BBN* [39]	–	79.82	82.18	–	88.32
L2RW* [7]	66.51	74.16	78.93	82.12	85.19
HCKC* [12]	–	77.05	–	–	87.51
Meta-weight net* [24]	68.91	75.21	80.06	84.94	87.84
M2m-LDAM* [19]	–	79.10	–	–	87.50
MetaSAug with CE [33]	75.95	79.84	84.02	87.07	88.42
PLA with CE* [26]	–	73.72	78.23	–	88.46
DPA-EI (Ours)	**77.82**	**82.02**	**84.22**	**87.16**	**88.75**

Table 5
Test accuracy (%) of ResNet-32 on CIFAR-LT-100 with different imbalance ratios. Rows with "*" indicate results reported in the original paper.

Imbalance ratio (ρ)	200	100	50	20	10
CE [34]	35.37	38.59	44.28	51.46	55.80
CB-CE* [21]	36.23	39.60	45.32	52.59	57.99
Focal* [22]	35.62	38.41	44.32	51.95	55.78
BBN* [39]	–	42.56	47.02	–	59.12
L2RW* [7]	33.38	40.23	44.44	51.64	53.73
HCKC [12]	35.15	38.06	43.07	51.49	56.80
CMO* [13]	–	41.58	47.26	–	59.16
DP-DPM* [40]	40.53	44.79	49.32	–	59.47
Meta-class-weight* [23]	39.31	43.35	48.53	55.62	59.58
Meta-weight-net* [24]	37.91	42.09	46.74	54.37	58.46
MetaSAug with CE [33]	39.13	44.71	49.52	55.25	59.72
DODA with CE* [29]	–	44.50	48.00	–	59.90
PLA with CE* [26]	–	41.77	46.34	–	59.10
ECS-SC with CE* [41]	–	43.16	47.32	–	59.68
DPA-EI (Ours)	**41.60**	**46.81**	**51.57**	**56.31**	**59.98**

Table 6
Test accuracy (%) on ImageNet-LT. "†" and "*" denote results borrowed from [41] and original paper, respectively.

Method	Many	Med	Few	All
CE† [34]	64.00	33.80	5.80	41.60
CB-Focal [21]	39.60	32.70	16.80	33.20
Focal† [22]	36.40	29.90	16.00	30.50
LDAM† [38]	**68.45**	41.76	13.42	48.07
cRT† [43]	58.77	44.03	26.12	47.35
τ-normalized* [43]	56.60	44.20	27.40	46.70
MetaSAug with CE [33]	61.95	48.70	26.97	46.74
CMO† [13]	67.27	40.70	15.51	47.80
CUDA† [42]	67.20	47.00	13.50	47.32
DODA with CE* [29]	67.40	47.50	13.90	48.10
ECS-SC with CE* [41]	67.32	41.61	23.94	48.93
DPA-EI (Ours)	63.96	**49.63**	**28.87**	**49.02**

4.4. Comparison with long-tailed methods

Experiments on CIFAR-LT. We conduct comparative experiments on the CIFAR-LT-10 and CIFAR-LT-100. We use ResNet-32 [36] as the basic backbone and the Stochastic Gradient Decent (SGD) optimizer with a momentum of 0.9. We apply a weight decay of 1×10^{-4} and train the models on a single GPU for 200 epochs in all experiments. Additionally, we decay the learning rate by 0.01 at the 160th and 180th epochs. We initialize the learning rate at 0.1 and set the batch size to 100 as [23].

Table 4 presents the experimental results for CIFAR-LT-10 with various imbalance ratios. The key findings are as follows:

(1) Compared with basic loss function methods, including CE loss [34], CB loss [21], and Focal loss [22], DPA-EI consistently achieves superior performance. Notably, DPA-EI improves accuracy by 11.67% over CE loss on CIFAR-LT-10 ($\rho = 100$), highlighting its effectiveness in addressing the imbalance in long-tailed classification tasks.

(2) DPA-EI achieves optimal accuracy under highly imbalanced conditions. For instance, it significantly outperforms other methods when $\rho = \{50, 100, 200\}$. Moreover, DPA-EI maintains a high level of accuracy even under less imbalanced conditions ($\rho = 20$ and $\rho = 10$). These results demonstrate the effectiveness and robustness of semantic augmentation in addressing long-tailed classification challenges.

Table 5 presents the classification accuracy of CIFAR-LT-100. The following observations are obtained:

(1) CB-CE [21] outperforms CE [34], indicating that the re-weighting strategy is effective in long-tailed scenarios. Accordingly, DPA-EI incorporates a re-weighting strategy to enhance the semantic augmentation process. Additionally, DPA-EI outperforms recent meta-learning methods, including L2RW [7], Meta-class-weight [23], and Meta-weight-net [24]. Notably, DPA-EI exhibits significant advantages on more imbalanced datasets.

(2) The results demonstrate that DPA-EI significantly outperforms both pure explicit and implicit data augmentation methods, indicating that a hybrid augmentation strategy is better suited for addressing long-tailed classification challenges. Notably, DPA-EI outperforms the best competitive method, MetaSAug with CE, by 2.47% under highly imbalanced conditions ($\rho = 200$).

Experiments on ImageNet-LT. We employ ResNet-50 [36] as the basic network for experiments on the ImageNet-LT dataset. The ResNet-50 is trained by SGD for 110 epochs with a momentum of 0.9 and a weight decay 3×10^{-4} as [23]. The experiment results on ImageNet-LT are shown in Table 6. This dataset evaluates many-shot, medium-shot, few-shot, and all-shot performance. From Table 6, we draw the following observations:

(1) Compared with CE [34], most methods lose some accuracy under the many-shot. Only a few methods simultaneously improve the accuracy of many, medium, and few shots. DPA-EI is slightly lower than CE by 0.04% under many-shot. In contrast, DPA-EI exceeds CE by 15.83% for medium-shot and 23.07% for few-shot. That means DPA-EI significantly facilitates learning for tail categories while causing negligible damage to head categories.

(2) Due to significant superiority under the few-shot and medium-shot, as well as the comparable performance under the many-shot, the overall performance of DPA-EI accuracy attained 49.02%. Specifically, DPA-EI surpasses the second-best method MetaSAug [33] by 0.93% accuracy under medium-shot. Similarly, DPA-EI outperforms the second-best method τ-normalized [43] by 1.47% under few-shot.

Experiments on iNaturalist. We apply ResNet-50 [36] as the backbone network for experiments on the iNaturalist 2018 dataset. The ResNet-50 model is trained with an initial learning rate of 0.1 and a batch size of 64. Table 7 compares DPA-EI with other methods on the iNaturalist 2018 dataset.

Most comparison methods use CE loss as their foundational loss function. Compared with these methods, DPA-EI further enhances the performance of CE loss. Notably, CE loss exhibits the lowest performance because it does not address the challenges posed by long-tailed datasets. In contrast, DPA-EI outperforms all competing methods, demonstrating its effectiveness and potential for tackling long-tailed classification tasks.

4.5. Analysis of dual progressive augmentation strategy

In this section, we evaluate the effectiveness of stage parameters in enhancing model classification performance. Specifically, we conduct five experiments, incrementally increasing the number of epochs by steps of 20. The experiments are conducted over a total of 200 epochs, with different stage configurations represented as 100_100, 120_80, 140_60, 180_20, and 200_0. In this format, "T_1_T_2" indicates that "T_1" corresponds to the number of epochs allocated for EDAS training, while "T_2" corresponds to the number of epochs allocated for IDAS

Table 7
Test accuracy (%) on iNaturalist 2018. "*" denotes results borrowed from the original paper.

Method	iNaturalist 2018
CE [34]	65.76
CB-CE* [21]	66.43
cRT* [43]	68.20
LDAM* [38]	64.58
LDAM-DRW* [38]	68.00
BBN* [39]	66.29
Meta-class-weight* [23]	67.55
MetaSAug with CE* [33]	68.75
CMO* [13]	68.90
DODA with CE* [29]	63.60
PLA with CE* [26]	66.70
DPA-EI (Ours)	**69.01**

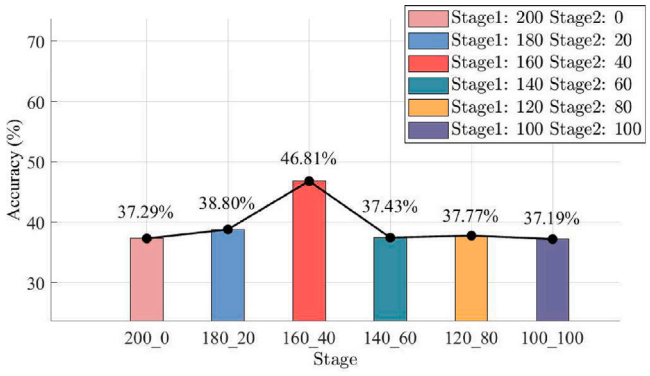


Fig. 5. Accuracy of different stage settings on CIFAR-LT-100 (ρ = 100).

training. Fig. 5 demonstrates the final accuracy comparisons under these different stage settings.

As shown in Fig. 5, the accuracy of DPA-EI reaches 46.81% with the 160_40 setting, improving performance by 8.01%~9.62% compared with other stage configurations. To better understand the impact of stage parameters on our experiment, we use the 160_40 setting as a boundary and further analyze the accuracy at each epoch during different stages. Data from the first stage, with fewer than 160 epochs, are presented in Fig. 6, while data from the second stage, with more than 160 epochs, are shown in Fig. 7. Additionally, we include 200 epochs of EDAS experimental data for comparison in each figure.

As shown in Fig. 6, we observe a significant improvement in the accuracy curve with the 160_40 setting. In contrast, the accuracy decreases in the corresponding epochs for the other experiments. For instance, the 100_100 setting shows a decline at 100 epochs, and similar decreases are observed at 120 and 140 epochs in the respective experiments. There is little difference in the final accuracy among the other experiments, except for the 160_40 setting. Compared with the experiments that apply 200 epochs of EDAS components individually, the experiments with fewer than 160 epochs exhibit minimal performance improvement, and in some cases, even slightly reduced performance. This suggests that explicit data augmentation requires a certain number of epochs to enhance the subsequent implicit data augmentation.

Ensuring a minimum of 160 epochs of EDAS training leads to superior experimental outcomes compared with using only the EDAS component for training. In Fig. 7, performance improves by 1.51% when the IDAS strategy is set to 20 epochs. However, increasing the number of IDAS epochs to 40 leads to an overall performance improvement of 9.25%. These results highlight the importance of providing sufficient training epochs for implicit data augmentation to fully realize its potential.

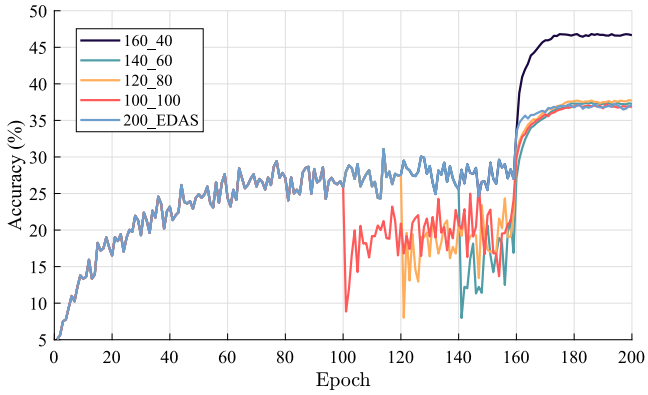


Fig. 6. Accuracy of different stage settings on CIFAR-LT-100 (ρ = 100). The first epoch parameter setting has fewer than 160 epochs.

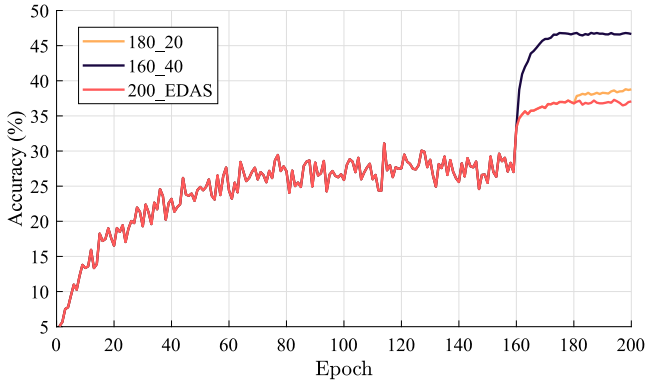


Fig. 7. Accuracy of different stage settings on CIFAR-LT-100 (ρ = 100). The first epoch parameter setting has more than 160 epochs.

4.6. Ablation study

Merit of each component. To investigate the impact of each component in DPA-EI, we conduct ablation experiments on the CIFAR-LT-100 dataset. The results are presented in Table 8, where "EDAS" and "IDAS" indicate whether the respective components are enabled. The findings in Table 8 demonstrate the following:

(1) Performance improvements are observed when EDAS and IDAS are applied individually. With EDAS, image-level augmentation methods enhance the generalization ability of the model. With IDAS, feature-level augmentation leads to significantly improved performance. Additionally, using EDAS for pre-training in the early epochs, followed by the IDAS strategy in the later epochs, significantly improves accuracy.

(2) Meta-learning plays a crucial role in IDAS, as its removal leads to a noticeable drop in performance. This demonstrates that the meta-learning algorithm learns the appropriate covariance, which is vital for classification. Therefore, this underscores the importance of our meta-learning approach in IDAS.

(3) The last row of the table shows the performance improvement achieved when both EDAS and IDAS are applied simultaneously. This combination leads to a significant performance boost, outperforming the use of any individual component. For example, when all components are employed together, performance improves by 4.18%, 7.29%, and 8.22% for ρ = {10, 50, 100}. This confirms that the dual data augmentation strategies of DPA-EI are progressive and have a positive impact on feature augmentation.

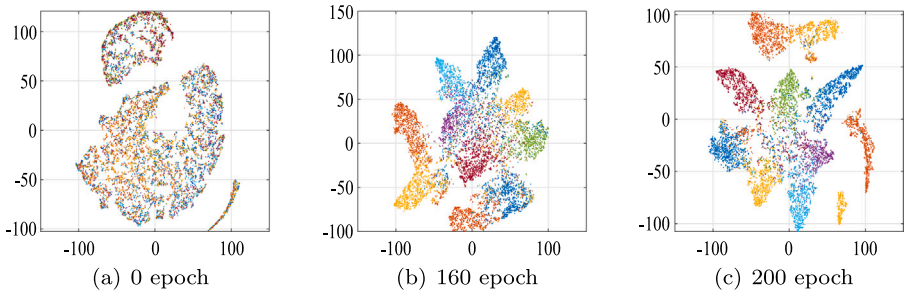


Fig. 8. T-SNE visualization at different epochs on CIFAR-LT-10 (ρ = 100). The 0 epoch represents the untrained model, while the 160 epoch indicates the use of only the EDAS strategy. The 200 epoch represents full training of the model, incorporating both EDAS and IDAS strategies, i.e., DPA-EI.

Table 8
Ablation study showing the accuracy (%) of each component on CIFAR-LT-100. The absolute improvements over the CE baseline are reported in brackets. "w/ meta-learning" represents IDAS with meta-learning, while "w/o meta-learning" represents IDAS without meta-learning.

EDAS	IDAS		Imbalance ratio (ρ)		
	w/ meta-learning	w/o meta-learning	100	50	10
			38.59	44.28	55.80
✓			39.54 (+0.95)	44.99 (+0.71)	58.02 (+2.22)
	✓		44.71 (+6.12)	49.52 (+5.24)	59.72 (+3.92)
✓		✓	45.41 (+6.82)	50.98 (+6.70)	59.58 (+3.78)
✓	✓		**46.81 (+8.22)**	**51.57 (+7.29)**	**59.98 (+4.18)**

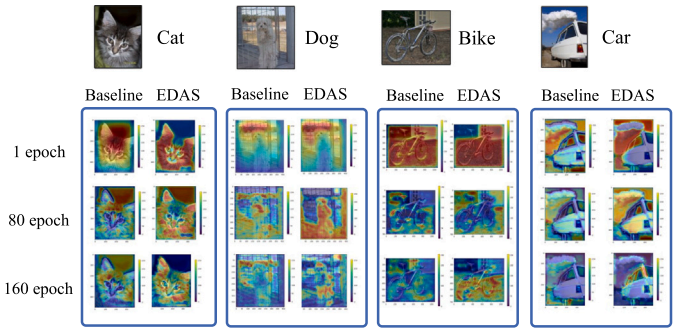


Fig. 9. Visualization of models attention for the four categories: *Cat*, *Dog*, *Bike*, and *Car*.

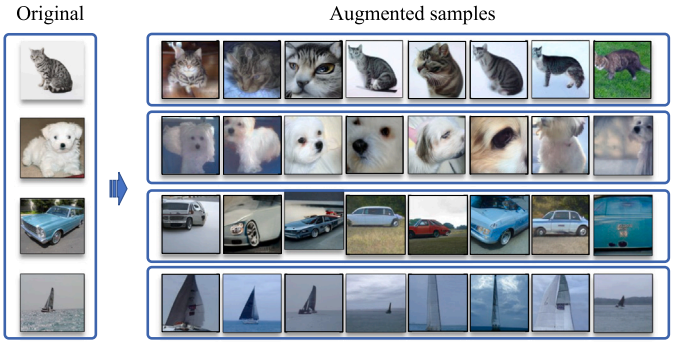


Fig. 10. Visualization of semantically augmented examples for the four categories: *Cat*, *Dog*, *Car*, and *Ship*.

4.7. Visualization results

Feature visualization. Fig. 8 shows the t-SNE visualization of CIFAR-LT-10 with ρ = 100, where different colors represent different categories. As shown in Fig. 8(a), the feature map appears chaotic and disorganized at the initial stage (0 epochs). This indicates that the features extracted by the untrained model lack clear category distinctions, leading to low accuracy. To compare the representations learned by our DPA-EI, we present t-SNE plots of feature changes at 160 and 200 epochs.

As shown in Fig. 8(b), the feature map begins to display a preliminary classification of categories, but there are still mixed representations and ambiguous category boundaries, indicating boundary distortion. In contrast, Fig. 8(c) shows clearer boundaries that separate different categories. This demonstrates the effectiveness of DPA-EI in gradually improving feature learning and enhancing category discrimination abilities during training.

Visual model attention of EDAS. Previous experiments have demonstrated that EDAS improves classification by enhancing the generalization ability of the model, as demonstrated in Figs. 8(a) and 8(b). To further analyze whether EDAS learns meaningful information from images, we use Grad-CAM [35] visualization, as shown in Fig. 9.

EDAS places greater emphasis on identifying and classifying key features compared with the baseline model. For example, when classifying a cat, EDAS focuses on the discriminative features rather than the majority of the image. In more complex scenarios, EDAS also highlights the main subject. Specifically, when analyzing a dog, the EDAS model focuses on the "body," while the baseline model focuses on the "forest" in the background. This ability of EDAS to concentrate on the relevant category features enhances the generalization ability of the model.

Visualization of augmented samples. To intuitively demonstrate that DPA-EI generates more diverse samples, we use the visualization method proposed in ISDA [14] to display some implicitly generated samples. Fig. 10 shows that DPA-EI produces samples with varying postures, angles, colors, and backgrounds. These changes are not derived from the training dataset itself but rather from samples within similar categories. DPA-EI semantically modifies the training examples while preserving their label identity.

5. Conclusions and future work

In this paper, we propose a framework for long-tailed classification called DPA-EI. DPA-EI balances sample-level generalization with feature-level semantic richness, avoiding issues such as boundary distortion and weak discrimination. We employ explicit data augmentation to enhance the samples, improving the generalization of the model. Furthermore, we introduce implicit augmentation to enrich the semantic content of the data by learning category-wise covariance matrices, addressing the challenges of diversity and representation in long-tailed classification. The experimental results demonstrate that DPA-EI achieves satisfactory performance on long-tailed datasets.

Overall, DPA-EI is a dual progressive augmentation framework that enhances both generalization and semantic richness. In the future, we aim to further integrate explicit and implicit augmentation methods, strengthening the progressive effect.

CRediT authorship contribution statement

Yan Zhao: Writing – original draft, Software, Methodology, Conceptualization. **Wenwei He:** Validation, Methodology, Investigation. **Hong Zhao:** Writing – review & editing, Validation, Methodology, Conceptualization.

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.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant No. 62376114 and the Fujian Province Undergraduate Universities Teaching Reform Research Project, China under Grant No. FBJY20240103.

Data availability

Data will be made available on request.

References

[1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. Berg, F. Li, Imagenet large scale visual recognition challenge, Int. J. Comput. Vis. 115 (3) (2015) 211–252.
[2] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C.L. Zitnick, Microsoft COCO: Common objects in context, in: European Conference on Computer Vision, 2014, pp. 740–755.
[3] Y. Kang, J. Zheng, M. Yang, N. An, Inter-structure and intra-semantics graph contrastive learning for disease prediction, Knowl.-Based Syst. (2024) 112059.
[4] X. Li, S. Liang, Y. Hou, T. Ma, StratMed: Relevance stratification between biomedical entities for sparsity on medication recommendation, Knowl.-Based Syst. 284 (2024) 111239.
[5] D. Liu, S. Zhong, L. Lin, M. Zhao, X. Fu, X. Liu, HOOST: A novel hyperplane-oriented over-sampling technique for imbalanced fault detection of aero-engines, Knowl.-Based Syst. 300 (2024) 112142.
[6] B. Krawczyk, M. Koziarski, M. Woźniak, Radial-based oversampling for multiclass imbalanced data classification, IEEE Trans. Neural Netw. Learn. Syst. 31 (8) (2019) 2818–2831.
[7] M. Ren, W. Zeng, B. Yang, R. Urtasun, Learning to reweight examples for robust deep learning, in: International Conference on Machine Learning, 2018, pp. 4334–4343.
[8] S. Park, J. Lim, Y. Jeon, J.Y. Choi, Influence-balanced loss for imbalanced visual classification, in: IEEE International Conference on Computer Vision, 2021, pp. 735–744.
[9] S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, Y. Yoo, Cutmix: Regularization strategy to train strong classifiers with localizable features, in: IEEE International Conference on Computer Vision, 2019, pp. 6023–6032.
[10] H. Zhang, M. Cisse, Y.N. Dauphin, D. LopezPaz, Mixup: Beyond empirical risk minimization, in: International Conference on Learning Representations, 2018, pp. 1–13.
[11] Y. Wang, E. Huang, R. Wang, T. Leng, BSDA in visual recognition: Balanced semantic data augmentation for long-tailed data, in: International Joint Conference on Neural Networks, 2023, pp. 1–8.
[12] H. Zhao, Z. Li, W. He, Y. Zhao, Hierarchical convolutional neural network with knowledge complementation for long-tailed classification, ACM Trans. Knowl. Discov. Data 18 (6) (2024) 1–22.
[13] S. Park, Y. Hong, B. Heo, S. Yun, J.Y. Choi, The majority can help the minority: Context-rich minority oversampling for long-tailed classification, in: IEEE Conference on Computer Vision and Pattern Recognition, 2022, pp. 6887–6896.
[14] Y. Wang, G. Huang, S. Song, X. Pan, Y. Xia, C. Wu, Regularizing deep networks with semantic data augmentation, IEEE Trans. Pattern Anal. Mach. Intell. 44 (7) (2021) 3733–3748.
[15] X. Chen, Y. Zhou, D. Wu, W. Zhang, Y. Zhou, B. Li, W. Wang, Imagine by reasoning: A reasoning-based implicit semantic data augmentation for long-tailed classification, in: AAAI Conference on Artificial Intelligence, 2022, pp. 356–364.
[16] W. Zhao, H. Zhao, Hierarchical long-tailed classification based on multi-granularity knowledge transfer driven by multi-scale feature fusion, Pattern Recognit. 145 (2024) 109842.
[17] J. Byrd, Z. Lipton, What is the effect of importance weighting in deep learning? in: International Conference on Machine Learning, 2019, pp. 872–881.
[18] D. Mahajan, R. Girshick, V. Ramanathan, K. He, M. Paluri, Y. Li, A. Bharambe, L. Van Der Maaten, Exploring the limits of weakly supervised pretraining, in: European Conference on Computer Vision, 2018, pp. 181–196.
[19] J. Kim, J. Jeong, J. Shin, M2m: Imbalanced classification via major-to-minor translation, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 13896–13905.
[20] Y. Wu, F. Min, B. Zhang, X. Wang, Long-tailed image recognition through balancing discriminant quality, Artif. Intell. Rev. 56 (2023) 833–856.
[21] Y. Cui, M. Jia, T.Y. Lin, Y. Song, S. Belongie, Class-balanced loss based on effective number of samples, in: IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 9268–9277.
[22] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, in: IEEE International Conference on Computer Vision, 2017, pp. 2980–2988.
[23] M.A. Jamal, M. Brown, M.H. Yang, L. Wang, B. Gong, Rethinking class-balanced methods for long-tailed visual recognition from a domain adaptation perspective, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 7610–7619.
[24] J. Shu, Q. Xie, L. Yi, Q. Zhao, S. Zhou, Z. Xu, D. Meng, Meta-weight-net: Learning an explicit mapping for sample weighting, in: Advances in Neural Information Processing Systems, 2019, pp. 1919–1930.
[25] H.P. Chou, S.C. Chang, J. Pan, W. Wei, D. Juan, Remix: Rebalanced mixup, in: European Conference on Computer Vision, 2020, pp. 95–110.
[26] J. Chen, B. Su, Instance-specific semantic augmentation for long-tailed image classification, IEEE Trans. Image Process. 33 (2024) 2544–2557.
[27] Z. Zhong, J. Cui, S. Liu, J. Jia, Improving calibration for long-tailed recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 16489–16498.
[28] H. Pan, Y. Guo, M. Yu, J. Chen, Enhanced long-tailed recognition with contrastive CutMix augmentation, IEEE Trans. Image Process. (2024) 4215–4230.
[29] B. Wang, P. Wang, W. Xu, X. Wang, Y. Zhang, K. Wang, Y. Wang, Kill two birds with one stone: Rethinking data augmentation for deep long-tailed learning, in: International Conference on Learning Representations, 2024, pp. 1–24.
[30] C. Zhang, B. Jiang, Z. Wang, J. Yang, Y. Lu, X. Wu, W. Sheng, Efficient multi-view semi-supervised feature selection, Inform. Sci. 649 (2023) 119675.
[31] S. An, E. Zhao, C. Wang, G. Guo, S. Zhao, P. Li, Relative fuzzy rough approximations for feature selection and classification, IEEE Trans. Cybern. 53 (4) (2021) 2200–2210.
[32] J. Liu, Y. Sun, C. Han, Z. Dou, W. Li, Deep representation learning on long-tailed data: A learnable embedding augmentation perspective, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 2970–2979.
[33] S. Li, K. Gong, C.H. Liu, Y. Wang, F. Qiao, X. Cheng, MetaSAug: Meta semantic augmentation for long-tailed visual recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2021, pp. 5212–5221.
[34] P. De Boer, D.P. Kroese, S. Mannor, R.Y. Rubinstein, A tutorial on the cross-entropy method, Ann. Oper. Res. 134 (1) (2005) 19–67.
[35] R.R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: IEEE International Conference on Computer Vision, 2017, pp. 618–626.
[36] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
[37] G. Van Horn, O. Mac Aodha, Y. Song, Y. Cui, C. Sun, A. Shepard, H. Adam, P. Perona, S. Belongie, The inaturalist species classification and detection dataset, in: IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 8769–8778.
[38] K. Cao, C. Wei, A. Gaidon, N. Arechiga, T. Ma, Learning imbalanced datasets with label-distribution-aware margin loss, in: Advances in Neural Information Processing Systems, 2019, pp. 1567–1578.
[39] B. Zhou, Q. Cui, X. Wei, Z. Chen, BBN: Bilateral-branch network with cumulative learning for long-tailed visual recognition, in: IEEE Conference on Computer Vision and Pattern Recognition, 2020, pp. 9719–9728.
[40] H. Zhang, L. Zhu, X. Wang, Y. Yang, Divide and retain: A dual-phase modeling for long-tailed visual recognition, IEEE Trans. Neural Netw. Learn. Syst. 35 (10) (2023) 13538–13549.
[41] W. He, J. Xu, J. Shi, H. Zhao, ECS-SC: Long-tailed classification via data augmentation based on easily confused sample selection and combination, Expert Syst. Appl. 246 (2024) 123138.
[42] S. Ahn, J. Ko, S.Y. Yun, CUDA: Curriculum of data augmentation for long-tailed recognition, in: International Conference on Learning Representations, 2023, pp. 1–16.
[43] B. Kang, S. Xie, M. Rohrbach, Z. Yan, A. Gordo, J. Feng, Y. Kalantidis, Decoupling representation and classifier for long-tailed recognition, in: International Conference on Learning Representations, 2019, pp. 1–16.